

Extending the reliability and applicability of B3LYP

Igor Ying Zhang, Jianming Wu and Xin Xu*

Received 11th January 2010, Accepted 2nd March 2010

First published as an Advance Article on the web 7th April 2010

DOI: 10.1039/c000677g

B3LYP is by far the most popular density functional in chemistry. Nevertheless, there is growing evidence, showing that B3LYP (1) degrades as the system becomes larger, (2) underestimates reaction barrier heights, (3) yields too low bond dissociation enthalpies, (4) gives improper isomer energy differences, and (5) fails to bind van der Waals systems, *etc.*

1. Introduction

Density functional theory (DFT) has become a common tool for first principles quantum chemical calculations of the electronic structure and properties of many molecular and solid systems.^{1–3} No other first principles methods can achieve comparable accuracy at the same low cost. Exact in principle,⁴ DFT replaces the conventional *ab initio* wavefunction, which depends on 3N spatial variables, by the electron density, which depends only on the three spatial variables, as a means to reach a solution to the Schrödinger equation. With the exact exchange–correlation functional, DFT could take into full account of all complex many-body effects. Unfortunately, the exact exchange–correlation functional is unknown, making it essential to pursue more and more accurate and reliable approximate functionals.

Various approximations to the exchange–correlation energy have been developed and tested in recent decades.^{5–26} A foundation of most approaches is the local density approximation (LDA^{5,6}), which is based on solutions of the uniform electron gas (UEG), using only density $\rho(r)$ at the position of evaluation. It is well documented that LDA yields results of good or moderate accuracy for such properties as lattice

constants, bulk moduli, equilibrium geometries, and vibrational frequencies.²⁷ However, LDA leads to bond energies and cohesive energies far too large, making it “not useful for thermochemistry”.²⁷

In addition to density, the generalized gradient approximation (GGA^{7–12}) also includes the first-order gradient of the density $\nabla\rho(r)$. The most popular GGA functionals include the B88⁷ exchange functional of Becke, which is often combined with the LYP⁸ correlation functional, due to Lee–Yang–Parr, and the “non-empirical” exchange–correlation functionals, PW91⁹ and PBE,¹⁰ proposed by Perdew and co-workers. These GGAs significantly reduce the overbinding tendency of LDA, but generally remain inadequate for the thermochemistry of molecules.²⁷

The so-called *meta*-GGA (*e.g.* TPSS,¹³ VSXC¹⁴), expands GGA to include further the kinetic energy density τ , and/or the Laplacian of the density $\nabla^2\rho(r)$. Various versions of *meta*-GGAs generally perform similarly to GGAs in applications. Nevertheless, the domain of *meta*-GGA may have not been thoroughly explored as it has for GGA.

A big step toward greater accuracy was the introduction of hybrid methods^{15–19} that include some amount of “exact exchange” on the basis of the adiabatic connection formula. Based on the number of occurrences of functional names in the journal titles and abstracts analyzed from the ISI Web of Science (2007), hybrid functional B3LYP^{7,8,17} is concluded to

State Key Laboratory of Physical Chemistry of Solid Surfaces,
College for Chemistry and Chemical Engineering, Xiamen University,
Xiamen 361005, China. E-mail: xinxu@xmu.edu.cn



Igor Ying Zhang



Jianming Wu



Xin Xu

Xin Xu is a Professor at Xiamen University, China. Mr Igor Ying Zhang is a PhD candidate with Prof. Xu and Dr Jianming Wu is a senior researcher in Prof. Xu's group. Mr. Zhang is a principle investigator for development of the XYG3 density functional. Dr Wu is responsible for designing neural-network correction to improve B3LYP energetics. The research interests of Prof. Xu's group involve developing and applying quantum chemical methods to problems ranging from surface chemistry, catalysis and chemical biology.

be the main working-horse in computational chemistry.²⁸ Indeed, B3LYP is by far the most popular density functional in chemistry. Despite the progress in the field, and the appearance of several new functionals every year, B3LYP continues to dominate the field, representing 80% of the total occurrences of density functionals in the literature, in the period 1990–2006.²⁸

On the other hand, it is now well-documented that B3LYP is not good for everything. Known problems include (1) accumulating errors on heats of formation as the size of the system is increased;^{29,30} (2) increasing errors on C–X bond energies with increased alkylation;^{31,32} (3) failures to give reliable energy ordering of isomers;³³ (4) underestimation of reaction barrier heights;^{34,35} and (5) breakdown in the description of van der Waals (vdW) interactions.³⁵ Other approximate functionals also share, to various extents, these shortcomings.^{34,35}

How can we go beyond B3LYP? Exciting achievements include GGAs (PBE,¹⁰ OLYP²¹), *meta*-GGAs (TPSS,¹³ VSXC,¹⁴), hybrids (PBE0,^{23,24} X3LYP,^{12,18} MCY,¹⁹ BMK,²⁵ M06 family¹⁵) and double hybrids (B2PLYP,²⁰ B2GP-PLYP²⁶), just to name a few. Our strategies fall into two categories. One is to design a systematic correction scheme on top of B3LYP,^{36–38} such that all B3LYP data, already and continuously built-up in the literature, can be used with higher accuracy and thus higher reliability at no extra cost as compared to B3LYP. The other is to develop new density functional.³⁹ The new functional should maintain the advantage of B3LYP, while surmount its known difficulties, leading to a general functional with more predictive power.

This feature article presents the effort made by our group.^{36–39} We will first recapitulate in brief the theory of our approaches, and will then demonstrate their performance, using well-established benchmarking datasets, for the prediction of heats of formation,^{29,40} heats of isomerization,³³ bond dissociation energies,³⁷ reaction barrier heights^{34,35} and non-bonded interaction energies.³⁵ We will summarize this article by pointing out the limitations of our present approaches and will outline the direction for future improvements.

2. Theory

2.1 Theoretical background

Kohn–Sham (KS⁴¹) method assumes a non-interacting N -electron system, which has density ρ_s , as constructed by the KS orbitals $\{\phi_i(r)\}$, the same as the original many-body system ρ .

$$\rho = \rho_s = \sum_{i=1}^N |\phi_i(r)|^2 \quad (1)$$

In the KS method, the total energy of a many-body system is expressed as

$$E[\rho] = T_s[\rho] + U[\rho] + V_{\text{ext}}[\rho] + E_{\text{xc}}[\rho] \quad (2)$$

where U is the classic Coulomb energy

$$U[\rho] = \iint d^3r' d^3r \frac{\rho(r')\rho(r)}{|r' - r|} \quad (3)$$

V_{ext} is the external potential energy from the nuclei, T_s is the kinetic energy of the non-interacting system expressed explicitly in terms of the KS orbitals. E_{xc} is the so-called exchange–correlation energy, which covers anything left to represent $E[\rho]$ in eqn (2).

The practical advantage of writing $E[\rho]$ in the form of eqn (2) is that the unknown functional $E_{\text{xc}}[\rho]$ is typically much smaller than the known terms T_s , U and V_{ext} . One can thus hope that reasonably simple approximations for $E_{\text{xc}}[\rho]$ provide useful results for $E[\rho]$. Indeed, there has been an evolution of successively better approximations to this functional, that has already provided quite good accuracy for many problems.^{5–26}

Conventionally, an approximate DFT functional $E_{\text{xc}}^{\text{DFT}}$ is decomposed into its exchange part and the correlation part:

$$E_{\text{xc}}^{\text{DFT}} = E_{\text{x}}^{\text{DFT}} + E_{\text{c}}^{\text{DFT}} \quad (4)$$

Indeed, exchange energy can be written explicitly in terms of the KS orbitals as

$$E_{\text{x}}^{\text{exact}} = -\frac{1}{2} \sum_{ij} \iint d^3r' d^3r \frac{\phi_j^*(r')\phi_i^*(r)\phi_j(r)\phi_i(r')}{|r' - r|} \quad (5)$$

Eqn (5) is exact if the KS orbitals give the true density. As it shares the same form as in the Hartree–Fock (HF) theory, which uses the HF orbitals, $E_{\text{x}}^{\text{exact}}$ is also frequently called E_{x}^{HF} . Nevertheless, it is common that $E_{\text{x}}^{\text{DFT}}[\rho]$ and $E_{\text{c}}^{\text{DFT}}[\rho]$ are developed jointly so that errors in one part can be compensated for by the other part. Simple replacement of $E_{\text{x}}^{\text{DFT}}$ in eqn (4) with $E_{\text{x}}^{\text{exact}}$ does not lead to good functional performance. It is worthy to point out that U as in eqn (3) includes electron self-interaction explicitly. Approximated $E_{\text{xc}}^{\text{DFT}}$ is unable to remove this self-interaction error completely, which has been related to many deficiencies of common DFT.^{19,42}

Most $E_{\text{xc}}^{\text{DFT}}$ take the form as

$$E_{\text{xc}}^{\text{DFT}} = \int d^3r \rho(r) \epsilon_{\text{xc}}^{\text{LDA}}(\rho) F(\rho, \nabla\rho, \nabla^2\rho, \text{and/or}, \tau) \quad (6)$$

where $\epsilon_{\text{xc}}^{\text{LDA}}$ is the exchange–correlation energy density under the UEG approximation, depending only on the density at the point of evaluation. F is the enhancement factor of varying sophistication, leading to functionals of GGA^{7,9,11,12,21} and *meta*-GGA.^{13–15,25} While the covalent bonding is a local phenomenon, the vdW interaction between distant systems is due to a very long-ranged correlation effect, quite different in form from that of the uniform-gas. It is therefore unsurprising that a local $E_{\text{xc}}^{\text{DFT}}$ misses the long-ranged tail of the vdW interaction (*e.g.* 1,4-interaction).^{27,43} It is further revealed that conventional DFT is unable to treat the medium-range electron correlation (*e.g.* 1,3-interaction) properly,⁴⁴ leading to the failure of stereoelectronic effects in saturated main group molecules.

2.2 A doubly hybrid density functional XYG3

The adiabatic connection formalism provides a rigorous way to define $E_{\text{xc}}[\rho]$.^{16,17,19,45–48} It assumes an adiabatic path between the fictitious non-interacting KS system ($\lambda = 0$) and the physical system ($\lambda = 1$) while holding the electron density

ρ fixed at its physical $\lambda = 1$ value for all λ of a family of partially interacting N-electron systems:

$$E_{xc}[\rho] = \int_0^1 U_{xc,\lambda}[\rho] d\lambda \quad (7)$$

$U_{xc,\lambda}$ is the potential energy of exchange–correlation at intermediate coupling strength λ . The only problem is that the exact integrand $U_{xc,\lambda}$ is unknown.

Becke first used this formalism as a practical tool for functional construction.^{16,17} He assumed a linear model¹⁶

$$U_{xc,\lambda} = a + b\lambda \quad (8)$$

By taking $U_{xc,\lambda=0} = E_x^{\text{exact}}$, and $U_{xc,\lambda=1} \approx U_{xc,\lambda=1}^{\text{LDA}} = 1$, one may have

$$a = E_x^{\text{exact}}, b = E_{xc}^{\text{LDA}} - E_x^{\text{exact}} \quad (9)$$

This gives Becke's half-and-half functional¹⁶

$$E_{xc}^{\text{H\&H}}[\rho] = \frac{1}{2}(E_x^{\text{exact}} + E_x^{\text{LDA}}) + \frac{1}{2}E_c^{\text{LDA}} \quad (10)$$

The key message from eqn (10) is that it introduces nonlocality by replacing some portion of the local exchange energy with the exact (HF-like) exchange energy. The popular Becke's three-parameter functional is an empirical modification of eqn (10).¹⁷

$$E_{xc}^{\text{B3LYP}}[\rho] = E_{xc}^{\text{SVWN}} + c_1(E_x^{\text{exact}} - E_x^{\text{S}}) + c_2\Delta E_x^{\text{B}} + c_3\Delta E_c^{\text{LYP}}, \quad (11)$$

where ΔE_x^{B} and ΔE_c^{LYP} are the gradient-containing correction terms to the LDA exchange–correlation, and $\{c_1 = 0.20, c_2 = 0.72, c_3 = 0.81\}$ are constants fitted against selected experimental thermochemical data of the G2 set.⁴⁰

Instead of using the end-point $U_{xc,\lambda=1}$, we pointed out that an alternative to fixing the $\{a, b\}$ parameters in eqn (8) is to use the initial slope ($U'_{xc,\lambda=0}$).³⁹ According to the Görling–Levy theory of coupling-constant perturbation expansion,⁴⁹ $U'_{xc,\lambda=0}$ is defined rigorously as the second-order correlation energy:

$$U'_{xc,\lambda=0} = \left. \frac{\partial U_{xc,\lambda}}{\partial \lambda} \right|_{\lambda=0} = 2E_c^{\text{GL2}} \quad (12)$$

This leads to:

$$b = 2E_c^{\text{GL2}}. \quad (13)$$

Eqn (9) and (13) give two choices of b , which we may combine using empirical parameters, $\{b_1, b_2\}$, to optimize the functional performance:

$$E_c = b_1E_c^{\text{GL2}} + b_2(E_{xc}^{\text{DFT}} - E_x^{\text{exact}}). \quad (14)$$

In principle, $E_{xc}^{\text{DFT}} \approx (E_{xc}^{\text{DFT}} - E_x^{\text{exact}})$ contains a complete description of correlation effects, so that the second term of eqn (14) may be interpreted as a way to extrapolate the second-order perturbation to infinite order. Hence, we propose to use an empirical formula of the form:³⁹

$$E_{xc}^{\text{XYG3}}[\rho] = E_{xc}^{\text{SVWN}} + d_1(E_x^{\text{exact}} - E_x^{\text{S}}) + d_2\Delta E_x^{\text{B}} + d_3(E_c^{\text{GL2}} - E_c^{\text{LYP}}) + \Delta E_c^{\text{LYP}}. \quad (15)$$

In comparison with B3LYP of eqn (11), XYG3 of eqn (15) is a double hybrid DFT that mixes some exact exchange into E_x^{DFT} which also introduces a certain portion of E_c^{GL2} into E_c^{DFT} . In practice, we may approximate E_c^{GL2} from the B3LYP

orbitals $\{\varphi_i\}$ with eigenvalues $\{\varepsilon_i\}$, where the subscripts (i, j) and (α, β) denote the occupied and unoccupied KS orbitals, respectively as:

$$E_c^{\text{GL2}} \approx \frac{1}{4} \sum_{ij} \sum_{\alpha\beta} \frac{|\langle \varphi_i \varphi_j | \hat{v}_{ee} | \varphi_\alpha \varphi_\beta \rangle|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_\alpha - \varepsilon_\beta} \quad (16)$$

Here \hat{v}_{ee} is the electron–electron repulsion operator and the singles contribution is omitted as Grimme did in his well-known double hybrid functional B2PLYP.²⁰ By fitting to the thermochemical data of the G3/99 set, the final three parameters in eqn (15) are determined empirically as $\{d_1 = 0.8033, d_2 = 0.2107, d_3 = 0.3211\}$.³⁹ Other systems, which are not included in the G3/99 thermochemical set, are used as independent test cases to validate the XYG3 functional.

2.3 Neural-network (NN) correction of B3LYP

The quest for ever improving functional is a never ending task. Nevertheless, any approximate functional inevitably contains some errors. This is understandable, as the knowledge of the exact exchange–correlation functional is equivalent to exactly solving the many-body Schrödinger equation. Thus designing simple yet efficient ways to correct the remaining errors of existing methods is appealing and has attracted much attention.^{50–57}

For the wavefunction based method, there exists a systematic way to pursue higher accuracy.^{58,59} The well established G2⁴⁰ and G3²⁹ theories aimed to reproduce effectively the quadratic configuration interaction QCISD(T, FC)/6-311+G(3df,2p) and QCISD(T, Full)/6-311+G(3d2f,2df,2p) energies, respectively, through the extrapolation in the one-particle and many particle spaces based on a series of calculations at a lower level. A higher level correction (HLC) procedure was further designed to compensate for the remaining deficiencies of the method based on the numbers of α and β valence electrons of the systems.^{29,40} For DFT methods, however, there does not exist such an extrapolation procedure for better description of correlation effect and the Gn-HLC-like procedure was found to not work well.²⁹

There have been many attempts, to try to make corrections on top of DFT predicted heats of formation ΔH_f^0 .^{36–38,50–57} These include parametrization of atomic energies,^{50–52} bond/group additivity corrections,^{53–55} corrections considering spin multiplicities and charges,^{53,55,56} and corrections based on neural-networks,^{36–38,57} etc. We proposed the X1 method using neural-networks to correct the B3LYP errors.^{36–38} The X1 method is based on the well established G3/99 set for heats of formation, plus 170 additional molecules (the X1-1 set) with a more diverse chemical environment up to 32 heavy atoms (n-C₃₂H₆₆).

Our NN adopts a three-layer architecture (see Fig. 1),^{36,38} which has an input layer consisting of inputs from the physical descriptors, a hidden layer containing five hidden neurons, and an output layer that outputs the corrected values for ΔH_f^0 . It employs a hyperbolic tangent sigmoid transfer function for all the hidden layer neuron models and the output neuron. $\{W_{xij}\}$ and $\{W_{yij}\}$ are sets of the connection weights, where $\{W_{xij}\}$ connects the input neurons and the hidden neurons,

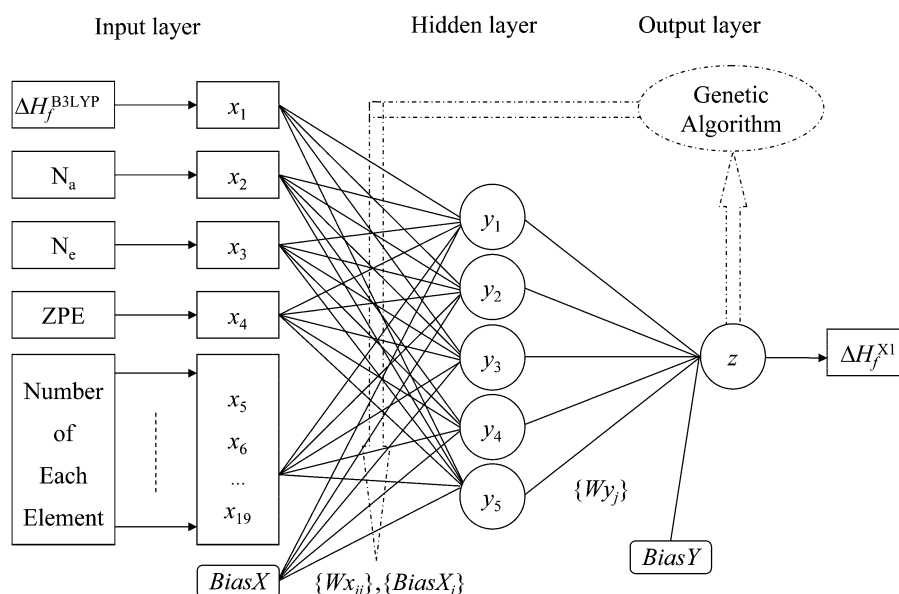


Fig. 1 Topological structure for neural-network correction to the B3LYP heats of formation.

and $\{W_{y_j}\}$ connects the hidden neurons and the output neuron.⁶⁰ The connection weights $\{W_{x_{ij}}\}$ and $\{W_{y_j}\}$ are optimized by using the general back propagation (BP) algorithm against the training set.^{60,61} To prevent optimization from trapping into a local minimum, a multi-population genetic algorithm (GA^{62,63}) is employed to cooperate with BP-NN.

The accuracy of NN depends critically on the choice of input descriptors. For generality and efficiency of X1, we chose $\Delta H_f^{\text{B3LYP}}$ (the B3LYP calculated heats of formation), N_a (the total number of atoms in a molecule), N_e (the total number of electrons in a molecule), ZPE (the calculated zero-point vibrational energy), and the number of each constituent elements (*e.g.* $N_H, N_C, N_N, N_O, N_F, N_{Si}, N_P, N_S, N_{Cl}$). Details on how to apply the NN correction may be found on the web site (<http://www.xdft.org/dft>).

For all the DFT calculations presented in this feature article, geometries are fully optimized by using B3LYP/6-311+G(d,p). Analytical vibrational frequencies are calculated at the same level. Single point calculations are performed with 6-311+G(3df,2p). B3LYP uses unscaled ZPE with no spin-orbit correction (SOC), which provides the raw data for NN corrections in X1.^{36,38} Scaled ZPE (0.9877) and SOC are included in XYG3.³⁹

3. Results and discussion

3.1 Heats of formation

Heats of formation, ΔH_f^0 , are among the most important chemical data. The stability of a molecule or the amount of energy released or absorbed in a reaction can be assessed based on the knowledge of accurate ΔH_f^0 . For stable molecules, ΔH_f^0 are typically obtained from calorimetric measurements.⁶⁴ This can be a tedious task and demands on using chemicals of very high purity. The situation is even less favorable for reactive intermediates such as free radicals, for which indirect methods have to be used and the results are frequently subjected to

substantial uncertainties.⁶⁵ Accurate computational chemistry methods to allow a reliable prediction of thermochemical data are therefore highly desirable.

We use the Gn paradigm developed by Pople and co-workers^{29,40} to show the performance of X1 and XYG3 to describe covalent bonding in the main group molecules. There are 148 molecules in the G2/97 set,⁴⁰ with an averaged number of non-hydrogen atoms of 2.6. Additional 75 molecules (the G3-3 set²⁹) have been added into G2/97, making a total of 223 molecules in the G3/99 set. The average number of non-hydrogen atoms in the G3-3 set is 5.8. The calculation of ΔH_f^0 is based on the theoretical atomization enthalpy of a molecule corrected by the experimental atomization enthalpies of the constituent elements in their standard states at 298 K.

Table 1 summarizes the statistical data for the predicted ΔH_f^0 of the G3/99 set. While B3LYP leads to a mean absolute deviation (MAD) of 5.6 kcal mol⁻¹, the X1 and XYG3 methods significantly improve it to 1.4³⁶ and 1.8³⁹ kcal mol⁻¹, respectively. These are comparable to the G2 theory (MAD = 1.9 kcal mol⁻¹), although still inferior to the G3 theory (MAD = 1.1 kcal mol⁻¹).^{36,39} The M06 family of functionals developed in Truhlar's group currently provides the highest accuracy with a broad applicability for chemistry.¹⁵ For the G3/99 set, M06-2x, M06 and M06-L lead to MADs of 2.9, 4.2 and 5.8 kcal mol⁻¹, respectively.³⁹

Table 1 also presents the results of B2PLYP, which is a pioneer doubly hybrid functional developed by Grimme.²⁰ With our present basis set (*i.e.*, 6-311+G(3df,2p) for electronic energy), B2PLYP gives MAD = 4.6 kcal mol⁻¹.³⁹ The salient difference between B2PLYP²⁰ and XYG3³⁹ is that B2PLYP employs the DFT portion of eqn (15) to generate the density used to calculate the DFT energy and orbitals from which the PT2 correction is computed. Such a truncated DFT may give density and orbitals that are dramatically different from the real ones.

We may further break down the 223 molecules of the G3/99 set into five general types of molecules, namely 48 non-hydrogen

Table 1 Statistic data (MAD/RMS)^a for the predicted heats of formation (298 K, kcal mol⁻¹) of the G3/99 set

Subsets ^b	B3LYP	X1	XYG3	B2PLYP	G2	G3
NH48	6.7/9.0 (22.2, SF ₆) ^c	1.8/2.4 (6.5, C ₂ F ₄)	3.7/5.1 (16.7, SF ₆)	3.4/4.4 (11.3, SF ₆)	3.0/3.9 (9.4, C ₂ F ₆)	2.1/2.8 (7.1, PF ₅)
HC38	7.6/9.4 (18.4, n-octane)	1.4/1.7 (3.5, bicyclobutane)	1.0/1.6 (5.7, naphthalene)	7.8/9.3 (20.4, n-octane)	1.7/2.8 (6.2, azulene)	0.7/0.9 (2.6, bicyclobutane)
SHC91	5.7/7.3 (16.9, tetrahydrothiopyran)	1.3/1.6 (4.9, acetylacetylene)	1.4/1.8 (6.6, dimethylsulfone)	5.3/6.8 (18.1, tetramethylsilane)	1.7/2.1 (4.5, acetic anhydride)	0.7/1.0 (3.6, vinyl chloride)
RD31	2.7/3.3 (8.0, BeH)	1.8/2.1 (4.1, phenyl)	1.1/1.4 (3.9, BeH)	2.0/2.9 (7.5, (CH ₃) ₃ C)	1.4/2.1 (7.1, phenyl)	0.8/1.0 (1.8, CN)
IH15	1.7/2.2 (5.5, N ₂ H ₄)	0.8/1.1 (2.3, H ₂ S)	2.1/2.9 (7.7, H ₂ O ₂)	2.0/2.4 (5.3, Si ₂ H ₆)	1.0/1.4 (2.9, Si ₂ H ₆)	0.9/1.1 (2.1, N ₂ H ₄)
G2/97 ^d	3.4/4.7	1.4/1.8	1.7/2.5	3.0/3.9	1.6/2.1	0.9/1.4
G3-3	9.9/11.0	1.6/2.0	2.1/3.4	7.8/9.3	2.5/3.3	1.3/1.9
Total	5.6/7.4	1.4/1.9	1.8/2.8	4.6/6.2	1.9/2.5	1.1/1.6

^a MAD: mean absolute deviation. RMS: Root-mean-squared errors. The experimental values are taken from ref. 29 and 40. ^b Codes for the subset names: NH = Nonhydrogens; HC = Hydrocarbons; SHC = Substituted Hydrocarbons; RD = Radicals; IH = Inorganic Hydrides. Followed with the set names are the numbers of molecules. ^c For each entry, the molecule which leads to the maximum absolute error is given in parentheses. ^d There are 148 molecules in the G2/97 set, and 75 molecules in the G3-3 set, making a total of 223 molecules in the G3 set.

systems (NH48), 38 hydrocarbons (HC38), 91 substituted hydrocarbons (SHC91), 31 radicals (RD31), and 15 inorganic hydrides (IH15) (see Table 1). The deviations (Expt.^{29,40}–Theor.) for each subset are depicted in Fig. 2. From Table 1 and Fig. 2, it is clear that B3LYP performs better for IH15 and RD31 (MADs = 1.7–2.7 kcal mol⁻¹) than for NH48, HC38 and SHC91 (MADs = 5.7–7.6 kcal mol⁻¹). The largest error occurs at SF₆ (–22.2 kcal mol⁻¹) of NH48. The negative deviation suggests a large under-binding tendency of B3LYP. Indeed, hypervalent molecules are also problematic for XYG3 and G3, for which the largest errors happen at SF₆ (–16.7 kcal mol⁻¹ for XYG3 and –7.1 kcal mol⁻¹ for G3). X1 is most satisfactory for the NH48 subset. For example, errors associated with X1 are –3.0 and 2.0 kcal mol⁻¹ for PF₅ and SF₆, respectively.

Generally, X1 and XYG3 are more accurate than B3LYP. The largest improvement in accuracy occurs for HC38 for which the MAD is reduced by more than a factor of 5, from 7.6 (B3LYP) to 1.4 (X1) and 1.0 (XYG3) kcal mol⁻¹. Based on the G3/99 set, one finds that B3LYP works better for the unsaturated hydrocarbons (MAD = 5.0 kcal mol⁻¹) than for the saturated hydrocarbons (MAD = 8.1 kcal mol⁻¹); while for X1, the difference between saturated and unsaturated

hydrocarbons is small (MAD = 1.3 vs. 1.4 kcal mol⁻¹). XYG3 is very satisfactory for the saturated hydrocarbons (MAD = 0.5 kcal mol⁻¹). It degrades for unsaturated hydrocarbons (MAD = 1.5 kcal mol⁻¹) with its largest errors occurring for azulene (4.6) and naphthalene (5.7 kcal mol⁻¹).

Substituted hydrocarbons occupy a large portion of the G3/99 set, covering various kinds of compounds such as alcohols, ethers, aldehydes, ketones and other oxygen, nitrogen, and halogen compounds, *etc.* B3LYP leads to MAD = 5.7 kcal mol⁻¹ for the SHC91 set, while X1 and XYG3 are much improved with MADs being around 1.4 kcal mol⁻¹. The maximum error associated with B3LYP for C₅H₁₀S (tetrahydrothiopyran, –16.9 kcal mol⁻¹) is effectively reduced to 0.6 and –1.1 kcal mol⁻¹, respectively, for X1 and XYG3.

Table 1 infers that B2PLYP does not improve B3LYP much for the calculations of heats of formation. However, this is in part due to the basis set effect. The original B2PLYP²⁰ demands a quadruple-zeta quality basis set to give MAD = 2.5 kcal mol⁻¹ for the G3/99 set, whereas the calculations performed here are at the triple-zeta quality basis set.

Fig. 2 suggests that there is an increasing underbinding tendency, *i.e.*, increased negative errors, for the B3LYP method as the size of the molecule is increased, whereas X1 and XYG3

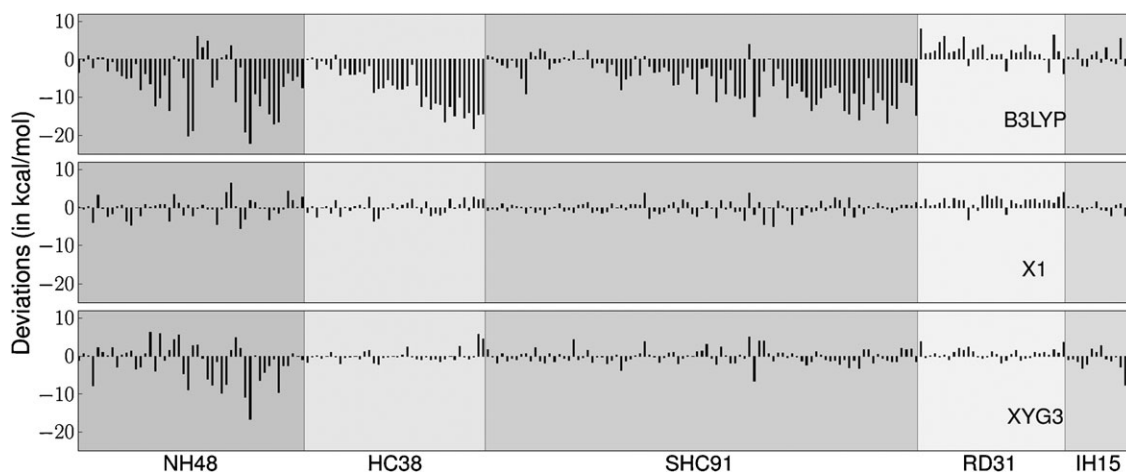


Fig. 2 Deviation (Expt.–Theor.) for theoretically predicted heats of formation. Abbreviations: NH = Nonhydrogens; HC = Hydrocarbons; SHC = Substituted Hydrocarbons; RD = Radicals; IH = Inorganic Hydrides. Followed with the set names are the numbers of molecules. Data are taken from ref. 38 and 39.

lead to a balanced error distribution. MAD of B3LYP increases from 3.4 for the G2/97 set to 9.9 kcal mol⁻¹ for the G3-3 set. Such a tendency is removed, to a large extent, in X1 (from 1.4 to 1.6 kcal mol⁻¹) and XYG3 (from 1.7 to 2.1 kcal mol⁻¹), whose quantities are similar to those of the Gn theory (see Table 1).

The significant size dependence is most evident by the calculated ΔH_f^0 of *n*-alkanes (see Fig. 3). For CH₄ and C₂H₆, B3LYP errors are 0.40 and -1.30 kcal mol⁻¹, respectively, which dramatically increase to -12.5 for *n*-C₆H₁₄, and -18.4 kcal mol⁻¹ for *n*-C₈H₁₈. B3LYP error is as high as -85.1 kcal mol⁻¹ for *n*-C₃₂H₆₆, suggesting that B3LYP should not be used for the prediction of ΔH_f^0 of medium and large size molecules. Fig. 3 shows that B2PLYP has similar errors as B3LYP for *n*-alkanes at the basis set of 6-311+G(3df,2p). Thus, B2PLYP leads to errors of -12.5 and -18.4 kcal mol⁻¹ for *n*-C₆H₁₄ and *n*-C₈H₁₈, respectively. If basis sets of quadruple-zeta quality are employed,²⁰ however, the corresponding errors are reduced to -7.3 (*n*-C₆H₁₄) and -10.4 (*n*-C₈H₁₈) kcal mol⁻¹. It is computationally demanding to directly calculate large molecules using the G3 method. Following Curtiss *et al.*,⁶⁶ we estimate the G3 heat of formation for *n*-C₃₂H₆₆ by using the CH₂ increment from smaller alkanes. The estimated ΔH_f^0 differs from the experimental value by 6.5 kcal mol⁻¹. Hence, errors accumulate even for G3. X1 and XYG3 turn out to be efficient yet accurate for *n*-alkanes. The largest error occurs at *n*-C₉H₂₀ for X1 (3.0 kcal mol⁻¹), while the maximum deviation for XYG3 is only 1.3 kcal mol⁻¹ for *n*-C₃₂H₆₆. As the G3/99 set contains only *n*-alkanes up to C₈, data for larger *n*-alkanes (C₉-C₃₂) provide an independent test, showing that the accumulating errors on heats of formation with the increased size of the system for B3LYP^{29,30} have been effectively alleviated.

The accurate prediction of heats of formation is one of the central topics in computational chemistry. The Gn (*n* = 1-3) family of model chemistries^{29,30} represents one of the most successful methods to date. Nevertheless, these methods are based on coupled-cluster-type treatments [QCISD(T)]. Hence, they are very computational-resource-demanding and computational-time-consuming. X1 and XYG3 offer promising alternatives. Indeed, as all these methods are developed according to a different philosophy, their agreement with each other and disagreement with the experiment is a strong indication that the experimental data is questionable. Table 2

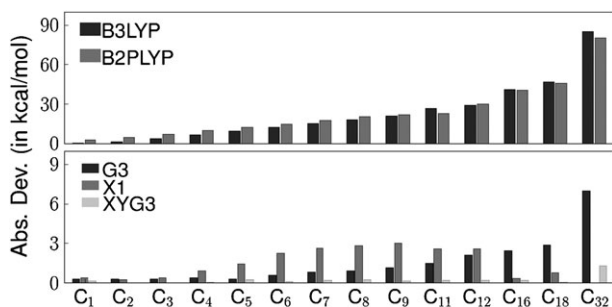


Fig. 3 Absolute deviations in calculated heats of formation for *n*-alkanes. G3 values for *n* > 11 are estimated by using CH₂ increments from smaller alkanes. Note the scale changes for the two panels.

summarizes several such examples. For compound (1), NIST WebBook⁶⁷ has listed three numbers, ranging from -77.7 to -82.2 kcal mol⁻¹. Even though the other two claim a small experimental uncertainty, all theoretical methods suggest that -82.2 ± 2.4 kcal mol⁻¹ is most reliable. This is also true for compound (2). The divergence for two reported experimental data is more than 15 kcal mol⁻¹. The theoretical methods all point to the low limit of 79.3 ± 2.3 kcal mol⁻¹. For compound (3), the calculated values are in the range from 22.2 to 24.5 kcal mol⁻¹, which do not agree with any of the two reported experimental data (9.94 or 36.0 kcal mol⁻¹). Thus the calculations suggest that both experimental data are likely incorrect. For compound (4), the underestimate tendency of B3LYP is clearly seen. Hence, the B3LYP predicted ΔH_f^0 is 13.8 kcal mol⁻¹ higher than the G3 value. The B2PLYP value is close to the B3LYP value, whereas the X1 and XYG3 numbers are close to the G3 number, which supports the experimental data -55.7 kcal mol⁻¹, used in the Gn theory,²⁹ and disproves the NIST value of -68.499 kcal mol⁻¹.⁶⁷ For compounds (5) and (6), it is too demanding, if not impossible, to carry out G3 calculations, whereas X1 and XYG3 can serve as an accurate yet efficient theoretical tool to detect the possible experimental errors. Indeed, X1 and XYG3 screen out one NIST value (-163.8 kcal mol⁻¹) as credible, and the other NIST value (-158.1 kcal mol⁻¹) as unlikely for compound (5). For compound (6), both X1 and XYG3 predict it as thermo-neutral (0.1-1.5 kcal mol⁻¹), which does not support the NIST value of -16.6 kcal mol⁻¹.

3.2 Bond dissociation enthalpies

Bond dissociation enthalpy (BDE) is defined here as the enthalpy change when a bond is cleaved by homolysis in the gas phase at 298 K and 1 atm: $X-Y(g) = X(g) + Y(g)$.⁶⁸ It is a central concept used everywhere in chemistry. Accurate BDE data are fundamental to the understanding of a diversity of chemical processes such as atmospheric and combustion reactions, or enzymatic catalysis, *etc.*⁶⁸⁻⁷⁰ Contrary to the general belief, we find that a good prediction of heats of formation (atomization energies) does not necessarily guarantee a good performance for BDE prediction.³⁷ If errors in heats of formation for radicals and the parent molecules are of opposite sign, there will be error accumulation in the prediction of bond dissociation energy. This is obviously the case for B3LYP. As shown in Fig. 2, there is a high frequency of negative errors for NH₄, HC₃₈ and SHC₉₁, whereas errors for RD₃₁ are generally positive, leading to accumulated errors in BDE predictions with B3LYP. Specifically, we have used 32 radicals and 116 molecules to set up 142 bond dissociation reactions.³⁷ B3LYP gives signed averaged deviations of -4.6 kcal mol⁻¹ for molecules, but 1.3 kcal mol⁻¹ for radicals, with a MAD = 4.7 kcal mol⁻¹ for all 148 ΔH_f^0 , which propagates into MAD = 6.3 kcal mol⁻¹ for 142 BDEs. For X1 and XYG3, such accumulative errors from ΔH_f^0 to BDE are largely amended. MADs for 148 ΔH_f^0 are around 1.4 kcal mol⁻¹ for X1 and XYG3, while those for 142 BDEs are 2.5 for X1 and 1.9 kcal mol⁻¹ for XYG3.

Fig. 4 depicts the error distribution for 142 BDEs broken down into five different types, including 16 C-H bonds,

Table 2 Prediction of heats of formation (kcal mol⁻¹). The best experimental values are highlighted in bold

	(1)	(2)	(3)	(4)	(5)	(6)
B3LYP	-83.5	72.3	23.6	-39.3	-137.0	23.3
X1	-83.6	76.9	24.5	-56.3	-162.6	0.1
XYG3	-82.8	76.8	22.9	-52.4	-165.2	1.5
B2PLYP	-84.7	74.4	24.3	-37.6	-151.5	23.1
G3	-84.5	76.2	22.2	-53.1	—	—
Expt.	-82.2 ± 2.4^a -77.7 ± 0.8 ^a -79.83 ± 0.20 ^a	79.3 ± 2.3^a 63.85 ^a	9.94^a 36.0 ^b	-55.7^c -68.499 ^a	-163.8 ± 3.1^a -158.1 ^a	-16.6 ± 2.1^a

^a Taken from ref. 67. ^b Taken from ref. 52. ^c Taken from ref. 29.

44 C–C bonds 14 X–Y bonds, 56 C–X bonds and 12 X–H bonds, where (X, Y = N, O, F, Si, P, S, and Cl). X1 and XYG3 are more accurate than B3LYP for all five types of bonds. The largest improvements occur at CX56/CC44, where MADs are reduced from 6.7/8.1 (B3LYP) to 2.4/2.9 (X1) and 1.8/1.7 kcal mol⁻¹ (XYG3).

From Fig. 4, it is obvious that B3LYP gives BDEs that are usually too small (*i.e.*, positive deviations) as compared to the experimental values. There are a few cases where BDEs are overestimated by B3LYP: deviations are -2.7 for H–CN, -2.0 for Cl–CN, -5.7 for NC–CN, -5.6 for H–CO, and -2.6 kcal mol⁻¹ for CH₃–CO. All of them involve formation of triple bond species, CO and CN, upon bond breaking. For X1/XYG3 methods, the errors are -2.3/0.7 for H–CN, -2.7/1.0 for Cl–CN, -5.1/0.4 for NC–CN, -3.0/-1.1 for H–CO, and -2.7/-0.7 kcal mol⁻¹ for CH₃–CO. Hence, X1 does not improve over B3LYP for these cases, whereas XYG3 is generally satisfactory.

Fig. 4 shows that there are some cases where XYG3 errors are significant (> 5 kcal mol⁻¹). They all involve NO₃ species (*e.g.* errors are 5.5 for *n*-C₃H₇–ONO₂, 6.2 for *i*-C₃H₇–ONO₂ and 5.9 kcal mol⁻¹ for C₂H₅–ONO₂). This radical presents a

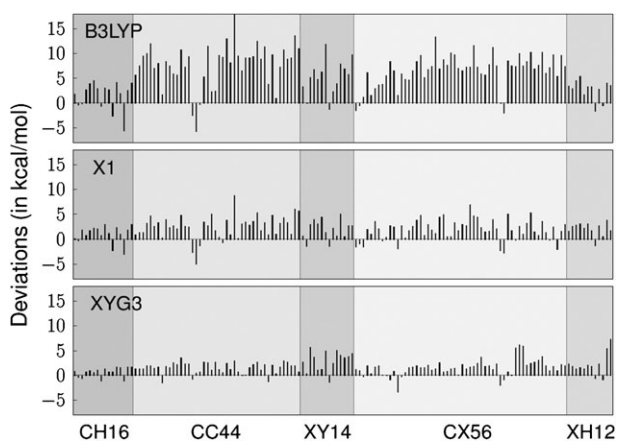


Fig. 4 Deviations (Expt.–Theor.) in calculated bond dissociation energies (X, Y = N, O, F, Si, P, S, Cl). Data are taken from ref. 37.

great challenge that different methods lead to substantially different results for ΔH_f° of NO₃ (*i.e.*, G3 21.8; XYG3 13.9; X1 14.4; B3LYP 8.5; B2PLYP 6.8 and M06-2x 26.5 kcal mol⁻¹). The experimental value is 17.0 kcal mol⁻¹.⁶⁷

Check and Gilbert have investigated the cumulative effect of the errors in large molecules by looking at the C–C bond-breaking energies of methyl-substituted ethanes.³¹ Indeed, even for simple C–C bonds in *n*-alkanes (CH₃–CH₃, CH₃–C₂H₅, C₂H₅–C₂H₅), B3LYP errors are as high as 5.7, 7.5 and 9.5 kcal mol⁻¹, respectively,³⁷ increasing with the number of carbon atoms (see Table 3). Errors are significantly increased when the carbon is highly alkylated.^{31,32} In (CH₃)₃C–C(CH₃)₃, B3LYP error for BDE has reached 21.6 kcal mol⁻¹.³⁷ The X1 method has removed, to a great extent, the errors associated with molecular size (*e.g.* Fig. 3), it nevertheless inherits some shortcomings of its parent B3LYP method. X1 error for BDE of (CH₃)₃C–C(CH₃)₃ is still as high as 11.8 kcal mol⁻¹.³⁷ Such kinds of problems have been overcome by XYG3. Inclusion of the PT2 term leads to substantial improvement on DFT correlation, giving BDE error for (CH₃)₃C–C(CH₃)₃ of only 2.8 kcal mol⁻¹, being competitive with G3MP2⁷¹ (*cf.* Table 3).

Check and Gilbert have also investigated the systematic underestimation of reaction energies for B3LYP as the number of C–C bonds increases by examining the progressive insertions of triplet methylene into C–H bonds of ethane.³¹ This is challenging as shown in Table 4. MAD of B3LYP for this set of 9 insertion reactions is as high as 18.6 kcal mol⁻¹. Even for G3MP2, MAD has reached 4.5 kcal mol⁻¹. X1 amends largely the B3LYP error, but remains inadequate for highly alkylated alkanes.³⁷ XYG3 is most satisfactory in this aspect (see Table 4).

3.3 Heats of isomerization

Accurate prediction of heats of isomerization is very important in organic chemistry. It is intertwined with quantifying some widely-used concepts such as conjugation, hyperconjugation, protobranching and resonance, *etc.*^{72–74} It also has important implications in many other fields such as life sciences and

Table 3 Bond dissociation enthalpies (298 K, kcal mol⁻¹) of internal C–C bond in progressively methyl-substituted alkanes

No.	Reaction	Expt. ^a	G3MP2 ^b	B3LYP ^b	X1 ^b	XYG3 ^b	B2PLYP ^b
1	CH ₃ –CH ₃ → CH ₃ + CH ₃	90.1	88.5 (1.6)	84.4 (5.7)	89.1 (1.0)	88.8 (1.3)	86.9 (3.2)
2	CH ₃ CH ₂ –CH ₃ → CH ₃ CH ₂ + CH ₃	88.9	88.2 (0.7)	81.4 (7.5)	87.5 (1.4)	87.6 (1.3)	85.0 (3.9)
3	(CH ₃) ₂ HC–CH ₃ → (CH ₃) ₂ HC + CH ₃	88.6	88.0 (0.6)	78.5 (10.1)	85.4 (3.2)	86.6 (2.0)	83.3 (5.3)
4	CH ₃ CH ₂ –CH ₂ CH ₃ → CH ₃ CH ₂ + CH ₂ CH ₃	87.8	88.0 (–0.2)	78.3 (9.5)	86.5 (1.3)	86.4 (1.4)	83.1 (4.7)
5	(CH ₃) ₃ C–CH ₃ → (CH ₃) ₃ C + CH ₃	87.5	88.0 (–0.5)	75.5 (12.0)	82.8 (4.7)	85.6 (1.9)	81.5 (6.0)
6	(CH ₃) ₂ HC–CH ₂ CH ₃ → (CH ₃) ₂ HC + CH ₂ CH ₃	87.3	86.5 (0.8)	74.4 (12.9)	83.5 (3.8)	84.8 (2.5)	80.6 (6.8)
7	(CH ₃) ₃ C–CH ₂ CH ₃ → (CH ₃) ₃ C + CH ₂ CH ₃	85.6	86.6 (–1.0)	70.5 (15.2)	79.6 (6.0)	83.2 (2.5)	78.0 (7.6)
8	(CH ₃) ₂ HC–CH(CH ₃) ₂ → (CH ₃) ₂ HC + CH(CH ₃) ₂	85.5	85.8 (–0.3)	69.1 (16.4)	79.3 (6.3)	82.2 (3.4)	76.9 (8.6)
9	(CH ₃) ₃ C–CH(CH ₃) ₂ → (CH ₃) ₃ C + CH(CH ₃) ₂	82.7	84.9 (–2.2)	64.0 (18.7)	73.9 (8.8)	79.8 (2.9)	73.4 (9.3)
10	(CH ₃) ₃ C–C(CH ₃) ₃ → (CH ₃) ₃ C + C(CH ₃) ₃	78.6	81.8 (–3.2)	57.0 (21.6)	66.8 (11.8)	75.8 (2.8)	68.2 (10.3)
	MAD/RMS^c	—	1.1/1.5	13.0/14.6	4.8/6.2	2.2/2.4	6.6/7.3

^a The experimental values are taken from ref. 67. ^b Theoretical errors in parentheses are given by (Expt. – Theor.). ^c MAD: mean absolute deviation. RMS: Root-mean-squared errors.

Table 4 Reaction enthalpies (298 K, kcal mol⁻¹) of progressive insertion of triplet CH₂ into the C–H bonds of ethane to form progressively methyl-substituted alkanes

No.	Reaction	Expt. ^a	G3MP2 ^b	B3LYP ^b	X1 ^b	XYG3 ^b	B2PLYP ^b
1	CH ₃ CH ₃ + 1 CH ₂ → CH ₃ CH ₂ CH ₃	–98.6	–97.3 (–1.3)	–94.2 (–4.4)	–98.3 (–0.3)	–98.4 (–0.2)	–96.4 (–2.2)
2	CH ₃ CH ₃ + 2 CH ₂ → (CH ₃) ₂ CHCH ₃	–199.4	–196.6 (–2.8)	–189.3 (–10.1)	–197.9 (–1.5)	–198.7 (–0.7)	–194.4 (–4.4)
3	CH ₃ CH ₃ + 2 CH ₂ → CH ₃ CH ₂ CH ₂ CH ₃	–197.3	–194.7 (–2.6)	–188.4 (–8.9)	–197.2 (–0.2)	–196.9 (–0.4)	–192.9 (–5.0)
4	CH ₃ CH ₃ + 3 CH ₂ → (CH ₃) ₃ CCH ₃	–301.2	–297.3 (–3.9)	–284.2 (–17.1)	–297.5 (–3.7)	–299.7 (–1.5)	–292.6 (–8.6)
5	CH ₃ CH ₃ + 3 CH ₂ → (CH ₃) ₂ CHCH ₂ CH ₃	–297.9	–292.7 (–5.3)	–282.4 (–15.4)	–295.8 (–2.1)	–296.5 (–1.4)	–290.0 (–7.9)
6	CH ₃ CH ₃ + 4 CH ₂ → (CH ₃) ₃ CCH ₂ CH ₃	–399.1	–393.4 (–5.7)	–376.3 (–22.8)	–394.1 (–5.0)	–396.9 (–2.2)	–387.47 (–11.7)
7	CH ₃ CH ₃ + 4 CH ₂ → (CH ₃) ₂ CHCH(CH ₃) ₂	–397.3	–391.5 (–5.8)	–375.2 (–22.1)	–393.3 (–3.9)	–395.1 (–2.2)	–386.0 (–11.3)
8	CH ₃ CH ₃ + 5 CH ₂ → (CH ₃) ₃ CCH(CH ₃) ₂	–497.3	–491.3 (–6.0)	–467.9 (–29.4)	–490.1 (–7.2)	–494.8 (–2.6)	–482.6 (–14.7)
9	CH ₃ CH ₃ + 6 CH ₂ → (CH ₃) ₃ CC(CH ₃) ₃	–596.1	–588.9 (–7.2)	–558.6 (–37.5)	–585.2 (–10.9)	–592.8 (–3.3)	–577.5 (–18.6)
	MAD/RMS^c	—	4.5/5.2	18.6/22.4	3.9/5.4	1.6/2.0	9.4/11.2

^a The experimental values are taken from ref. 67. ^b Theoretical errors in parentheses are given by (Expt. – Theor.). ^c MAD: mean absolute deviation. RMS: Root-mean-squared errors.

interstellar sciences, *etc.* For example, a major concern for understanding the possible role of interstellar complex organic molecules is to explore the stability and transformation among these organic isomers. The most stable isomer may also be the most abundant one in space, which may ultimately be at the origin of life.⁷⁵

The energy difference of propyne *versus* allene is a pathological case sorted out by Woodcock *et al.*⁷⁶ (see Table 5). While the experiment suggests that propyne is more stable than allene by 1.3 kcal mol⁻¹, all established DFT methods give the opposite energy ordering with considerable error bars.⁷⁶ We have examined the performance of some new functionals. Indeed, we find that BMK²⁶ predicts erroneously that allene is 1.3 kcal mol⁻¹ more stable than propyne. M06-2x¹⁵ reduces this error to 0.3 kcal mol⁻¹, but still favors allene to propyne. As shown in Table 4, both B2PLYP²⁰ and XYG3³⁹ give the correct energy ordering. B2PLYP underestimates the relative stability of propyne by 1.0 kcal mol⁻¹, while XYG3 overestimates it by 1.3 kcal mol⁻¹. In Table 5, we have also included the third isomer of C₃H₄, cyclopropene. X1, XYG3 and G3 lead to $\Delta H_f^0 = 68.2$ – 68.5 kcal mol⁻¹, which are ~ 2 kcal mol⁻¹ higher than the experimental value of 66.2 kcal mol⁻¹.^{29,40} B2PLYP with large basis sets also predict ΔH_f^0 cyclopropene of 68.8 kcal mol⁻¹.²⁰ All these results cast some doubt on the accuracy of the experimental data.


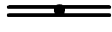

We notice that X1 is unable to correct the B3LYP error to give correct energy ordering between propyne and allene. The present version of X1 uses only atom types as descriptors and

thus can be generally applied easily. Discrimination of isomers in X1 relies only on B3LYP calculated ΔH_f^0 and ZPE. Such descriptors are not good enough for higher accuracy of isomerization energy.^{36,38} Inclusion of new descriptors of bond and group types is anticipated and work along this line is in progress.

Data of C₈H₁₀ presented in Table 6 provides a more thorough examination of theoretically predicted isomer stabilities.³³ As is clearly seen, directly calculated B3LYP ΔH_f^0 from atomization energies consistently underestimate the stabilities with a huge MAD₁ of 17.3 kcal mol⁻¹. Such a tendency is also obvious with B2PLYP, but we expect that large basis sets can amend, to a certain extent, the errors. X1 removes most of B3LYP errors, giving MAD₁ = 3.3 kcal mol⁻¹. G3 and XYG3 are most satisfactory, leading to MAD₁ of 1.4 and 1.6 kcal mol⁻¹, respectively. X1 is more accurate for alkylbenzenes (Nos. 1–3), but still has significant errors for ring-strained molecules (Nos. 6, 10, and 11). XYG3, on the other hand, performs best for ring-strained molecules, but is less satisfactory for alkylbenzenes. If NN correction is applied to XYG3, the results would be most satisfactory.















It is claimed that computing isomer energy differences is an easier task than estimating heats of formation.³³ This may be too optimistic. We examined the error of the energy differences for each pair of isomers. B3LYP and B2PLYP lead to MAD₂ of 6.1 and 3.5 kcal mol⁻¹, respectively, for energy differences, showing an improvement over MAD₁ of ΔH_f^0 . There is a degradation for G3, XYG3 and X1. Especially, MAD₂ of X1

Table 5 Heats of formation (298 K, kcal mol⁻¹) for C₃H₄ isomers.

Method	ΔH_f^\ominus			$\Delta\Delta H_f^\ominus$		
	 (1)	 (2)	 (3)	1-2	1-3	2-3
Expt. ^a	44.2	45.5	66.2	-1.3	-22.0	-20.7
G3 ^b	44.4 (-0.2)	45.0 (0.5)	68.4 (-2.2)	-0.6 (-0.7)	-24.0 (2.0)	-23.4 (2.7)
B3LYP ^b	46.8 (-2.6)	44.4 (1.1)	70.4 (-4.2)	2.4 (-3.7)	-23.6 (1.6)	-26.0 (5.3)
X1 ^b	45.7 (-1.5)	43.6 (1.9)	68.5 (-2.3)	2.1 (-3.4)	-22.8 (0.8)	-24.9 (4.2)
XYG3 ^b	43.1 (1.1)	45.7 (-0.2)	68.2 (-2.0)	-2.6 (1.3)	-25.1 (3.1)	-22.5 (1.8)
B2PLYP ^b	46.7 (-2.5)	47.0 (-1.5)	70.9 (-4.7)	-0.3 (-1.0)	-24.2 (2.2)	-23.9 (3.2)
B2PLYP ^{b,c}	44.4 (-0.2)	44.8 (0.7)	68.8 (-2.6)	-0.4 (-0.9)	-24.4 (2.4)	-24.0 (3.3)

^a The experimental values are taken from ref. 67. ^b Theoretical errors in parentheses are given by (Expt. - Theor.). ^c From ref. 20 using basis sets of quadruple-zeta quality.

Table 6 Heats of formation (298 K, kcal mol⁻¹) for C₈H₁₀ isomers

No.	Structures	Expt. ^a	G3 ^b	B3LYP ^b	X1 ^b	XYG3 ^b	B2PLYP ^b
1		4.1	3.8 (0.3)	17.4 (-13.3)	3.9 (0.3)	1.6 (2.5)	13.0 (-8.9)
2		4.3	4.0 (0.3)	16.9 (-12.6)	3.5 (0.8)	1.2 (3.1)	12.5 (-8.2)
3		4.5	4.3 (0.3)	18.0 (-13.4)	4.0 (0.6)	1.6 (3.0)	13.2 (-8.7)
4		34.3	35.0 (-0.7)	49.4 (-15.1)	35.7 (-1.4)	33.5 (0.8)	45.0 (-10.7)
5		37.7	35.4 (2.3)	58.7 (-21.0)	42.3 (-4.6)	36.1 (1.6)	50.9 (-13.2)
6		43.0	42.7 (0.3)	66.9 (-23.9)	49.8 (-6.8)	42.1 (0.9)	57.6 (-14.6)
7		45.2	46.8 (-1.6)	65.6 (-20.3)	49.5 (-4.3)	47.1 (-1.9)	60.1 (-14.8)
8		46.9	49.7 (-2.8)	66.1 (-19.2)	50.6 (-3.7)	50.0 (-3.1)	61.9 (-15.0)
9		53.1	53.7 (-0.6)	70.4 (-17.3)	54.8 (-1.7)	53.8 (-0.7)	65.9 (-12.8)
10		55.1	54.1 (1.0)	80.0 (-24.9)	62.8 (-7.7)	54.9 (0.2)	70.4 (-15.3)
11		57.1	56.1 (1.0)	82.1 (-25.0)	65.0 (-7.9)	57.0 (0.1)	72.4 (-15.3)
12		61.8	60.7 (1.1)	72.7 (-10.9)	64.9 (-3.1)	60.2 (1.6)	70.4 (-8.6)
13		62.5	67.7 (-5.2)	72.8 (-10.3)	59.2 (3.3)	60.2 (2.3)	70.2 (-7.7)
14		75.4	73.7 (1.7)	90.1 (-14.7)	75.0 (0.4)	74.4 (1.0)	85.9 (-10.5)
	MAD_1/RMS_1^c		1.4/2.0	17.3/18.7	3.3/4.4	1.6/2.0	11.7/12.5
	MAD_2/RMS_2^d		2.1/2.8	6.1/7.3	4.1/4.9	2.1/2.5	3.5/4.3

^a The experimental values are taken from ref. 67. ^b Theoretical errors in parentheses are given by (Expt. - Theor.). ^c Mean absolute deviation and root mean square error for calculated heats of formation. ^d Mean absolute deviation and root mean square error for calculated relative energies.

is as high as 4.1 kcal mol⁻¹. Even though there is some improvement over B3LYP, X1 is generally insufficient for the reliable description of isomerization energies. Design of more sophisticated NN descriptors is under way. XYG3, in comparison with G3 for cost-effectiveness, is a valuable tool for isomerization energy calculations.

3.4 Reaction barrier heights

Accurate prediction of reaction barrier heights is very important in detailed kinetic modeling of any chemical system such as fuel combustion,⁷⁷ catalytic processes,^{78,79} and chemical vapor deposition,^{80,81} *etc.* It is now well-known that various DFT methods generally underestimate the barrier heights.^{15,39} There are many investigations devoted to improve this situation. Outstandingly, Zhao *et al.* compiled several benchmark databases of barrier heights,^{15,34,35} which have been widely used to test existing functionals and/or to train new functionals. The so-called HTBH38/04 set includes forward and reverse barrier heights for 19 hydrogen transfer reactions; while the NHTBH38/04 set comprises forward and reverse barriers of 19 non-hydrogen transfer reactions, which are further partitioned into three subsets with (1) 6 heavy-atom transfer reactions, (2) 8 nucleophilic substitution reactions and

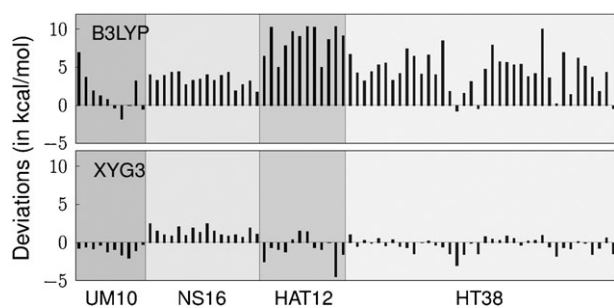


Fig. 5 Deviations (Ref.–Theor.) in calculated reaction barrier heights. Geometries and reference energies are taken from the database website of Truhlar's group.^{15,34,35} Abbreviations: UM = association and unimolecular reactions; NS = nucleophilic substitution reactions; HAT = heavy-atom transfer reactions; HT = hydrogen transfer reactions. Followed with the set names are the numbers of barriers.

(3) 5 association and unimolecular reactions. It was found³⁹ that errors associated with LDA for a total of 76 barrier heights of HTBH38/04 and NHTBH38/04 are MAD = 14.9 kcal mol⁻¹, while MADs for GGAs are 8.7 for PBE and 8.2 kcal mol⁻¹ for BLYP. *Meta*-GGAs may not represent an improvement over GGAs. Hence MAD for these two databases is as high as 8.3 kcal mol⁻¹ for TPSS. Hybrid GGA such as B3LYP offers a clear advantage, leading to MAD of 4.3 kcal mol⁻¹. Nevertheless, such an accuracy is not sufficiently good. Encouragingly, satisfactory results are obtained with the recent hybrid *meta*-GGAs: MAD = 2.1 (M06), and 1.3 kcal mol⁻¹ (M06-2x). MAD associated with doubly hybrid functional B2PLYP is 1.94 kcal mol⁻¹.³⁹

It is generally agreed that self-interaction errors (SIE) of local DFT functionals are responsible for the poor performance on the calculations of barrier heights.^{19,42} A large portion of exact exchange has been shown to be valuable. Indeed, B2PLYP and M06-2x contain ~53% E_x^{exact} , while that in XYG3 is as high as 80%, leading to MAD of 1.02 kcal mol⁻¹ for the total of 76 barrier heights. This is the same accuracy as would be obtained by using the QCISD(T) *ab initio* method with the same basis set.³⁹ We emphasize that barrier heights are *not* included in the B2PLYP and XYG3 training set, but are included in the M06 training set. B2PLYP and XYG3 are new generation functionals, which also include information of unoccupied orbitals.

Fig. 5 depicts the deviations (Ref.–Theor.) in calculated reaction barrier heights. As compared to the reference value computed with W1,^{15,34,35} the underestimating tendency (*i.e.*, positive error) of B3LYP is clearly seen. There are only a few exceptions where the deviation is negative. This is in line with the fact that B3LYP generally predicts too small BDE (see Fig. 4). Notably, B3LYP barrier for HCO → H + CO is overestimated by 1.8, while B3LYP BDE (H–CO) is too high by 5.6. The largest B3LYP errors (10.3 kcal mol⁻¹) occur at N₂ + OH → N₂O + H and HF + F → H + F₂. XYG3 reduces the error for the former reaction to 1.5 kcal mol⁻¹, while it still suffers from an error of –4.5 kcal mol⁻¹ for the second reaction. This is a reflection of a large error (7.8 kcal mol⁻¹) of XYG3 BDE (F–F). As shown in Fig. 5, HAT is most difficult for B3LYP. MAD for this subset

Table 7 Barrier heights (in kcal mol⁻¹) for several hydrogen transfer reactions: W1 reference data^a and errors (Ref.–Calc) for other methods

No.	Barrier heights	W1	B3LYP	X1	XYG3	B2PLYP
1	CH ₄ + H → CH ₃ + H ₂	15.3	5.7	1.3	0.3	1.6
2	CH ₃ + H ₂ → CH ₄ + H	12.1	3.2	0.2	0.3	2.2
3	H ₂ O + H → OH + H ₂	21.2	7.9	2.2	0.5	2.6
4	OH + H ₂ → H ₂ O + H	5.1	4.3	0.3	–0.5	1.7
5	CH ₄ + NH ₂ → CH ₃ + NH ₃	14.5	3.1	1.1	–0.1	1.6
6	CH ₃ + NH ₃ → CH ₄ + NH ₂	17.8	4.4	2.3	0.6	2.3
7	CH ₄ + OH → CH ₃ + H ₂ O	6.7	4.5	2.1	–0.1	2.2
8	CH ₃ + H ₂ O → CH ₄ + OH	19.6	5.6	2.9	0.9	2.6
9	NH ₃ + OH → NH ₂ + H ₂ O	3.2	5.6	2.3	–0.5	2.6
10	NH ₂ + H ₂ O → NH ₃ + OH	12.7	5.4	1.8	–0.4	2.2
11	H ₂ + O → H + OH	13.1	7.0	2.3	–0.7	3.1
12	H + OH → H ₂ + O	10.7	6.6	2.8	0.2	2.4
13	CH ₄ + NH → CH ₃ + NH ₂	22.4	5.2	2.1	–0.1	2.4
14	CH ₃ + NH ₂ → CH ₄ + NH	8.0	1.9	0.7	–1.5	0.1
	MAD^b		5.0	1.8	0.5	2.1

^a W1 reference data are taken from ref. 34 and 35. ^b Mean absolute deviation.

is 8.5 kcal mol⁻¹. B3LYP is most satisfactory for UM subset (MAD = 2.0 kcal mol⁻¹). XYG3 shows significant improvement over B3LYP. MAD for HAT is reduced to 1.38 kcal mol⁻¹, while that for UM is 0.98 kcal mol⁻¹.

The barrier height prediction should be correlated with the functional performance of BDE prediction. Furthermore, if unimolecular rearrangement transition structures are considered as isomers of ground-state molecules, then a method that has a better ability to reproduce isomer energy differences also implies its better ability for determining activation barriers of chemical reactions. As shown above, the X1 method can improve the B3LYP BDE prediction and isomer energy difference prediction, we suggest that X1 should also improve B3LYP over barrier height prediction. Table 7 summarizes the

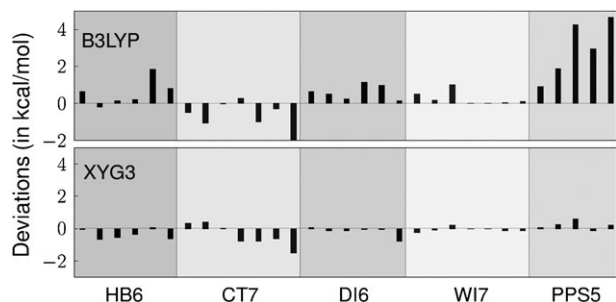
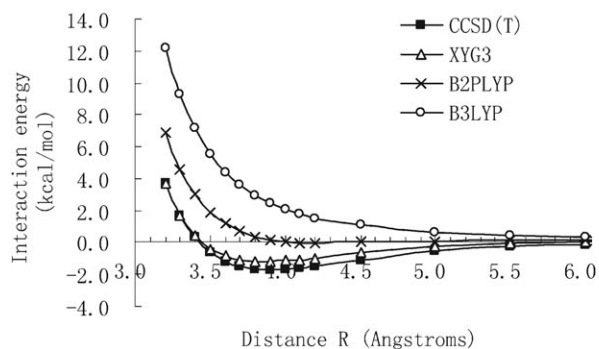
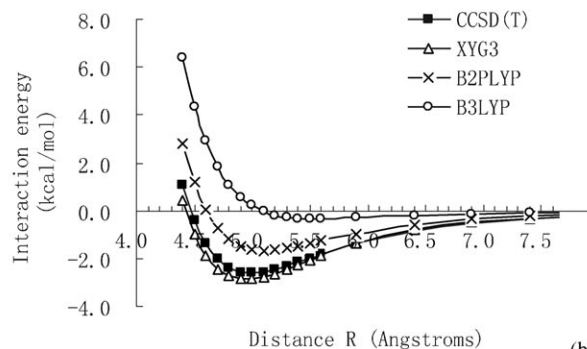


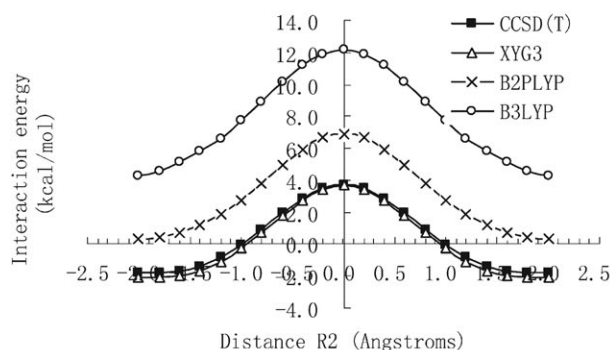
Fig. 6 Deviations (Ref.–Theor.) in calculated interaction energies of the NCIE31/05 set. Geometries and reference energies are taken from the database website of Truhlar's group.^{15,34} Abbreviations: HB = hydrogen bonded interactions; CT = charge transfer complexes; DI = dipole interactions; WI = weak interactions; PPS = π - π stacking complexes. Followed with the set names are the numbers of systems.



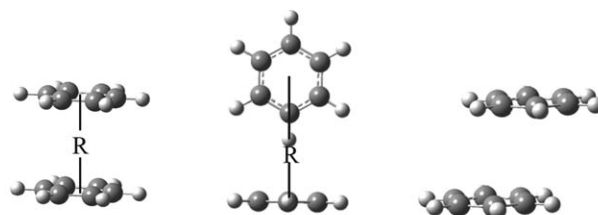
(a)



(b)



(c)



Sandwich (a)

T-shaped (b)

Parallel-displaced (c)

Fig. 7 Potential energy curves for sandwich (a), T-shaped (b) and parallel-displaced ($R_1 = 3.2$ Å) configurations of benzene dimer. CCSD(T) numbers are from ref. 93.

calculated barrier heights for 14 hydrogen transfer reactions, where Truhlar's W1 values^{34,35} are used as reference numbers. We re-optimize the geometries for transition states and local minimum at the level of B3LYP/6-311+G(d,p). As X1 is based on standard heats of formation at 298 K, we remove the thermo contributions $\Delta H_{0 \rightarrow 298}$ and the ZPE contributions calculated at B3LYP to get the classic barrier heights of X1. As shown in Table 7, B3LYP errors are between 1.9 and 7.9 kcal mol⁻¹, with MAD of 5.0 kcal mol⁻¹ for this set of barrier heights. X1 indeed improves over B3LYP, reducing the maximum error to 2.9 and MAD to 1.8 kcal mol⁻¹. Note that no barrier height data have been included for NN training.

3.5 Non-bonded interaction energies

Non-bonded interactions are undoubtedly important in many soft-matter situations such as supramolecular chemistry,⁸² protein folding⁸³ and polymer cohesion.⁸⁴ They also play an important role in the energetics of molecules' interaction with graphene sheets⁸⁵ or carbon nanotubes,⁸⁶ with potential implications for nanotechnology. Non-bonded interactions are usually determined by a complicated interplay between attractive and repulsive interactions. The possible attraction includes van der Waals (vdW or dispersion) of two nonpolar subsystems, polarization (or induction) of one nonpolar subsystem by a polar subsystem or electrostatic interaction between permanent multipoles.^{87–90} At shorter intermolecular distances, the dominant force is the exchange repulsion. It is now clear that common density functionals (*e.g.*, BLYP and PBE) cannot describe vdW interactions due to a very long-ranged correlation hole, that is quite different in form from the uniform-gas hole.⁴³ It is also clear that the exchange repulsion

Table 8 Heats of formation of CX₄ and SiX₄ (X = F and Cl), and the halogen exchange reaction enthalpy (kcal mol⁻¹)

	CF ₄	+	SiCl ₄	→	CCl ₄	+	SiF ₄	ΔH _r
B3LYP ^a	-218.9 (-4.2)		-139.7 (-18.8)		-9.4 (-13.5)		-365.7 (-20.3)	-16.6 (-10.9)
X1 ^a	-223.8 (0.8)		-156.2 (-2.2)		-19.3 (-3.6)		-386.7 (0.7)	-26.0 (-1.5)
XYG3 ^a	-221.9 (-1.2)		-161.2 (2.8)		-24.5 (1.5)		-377.2 (-8.8)	-18.6 (-8.9)
B2PLYP ^a	-225.3 (2.3)		-153.2 (-5.2)		-21.5 (-1.4)		-375.7 (-10.3)	-18.7 (-8.8)
G2 ^a	-228.6 (5.5)		-162.2 (3.8)		-25.8 (2.8)		-378.8 (-7.1)	-13.8 (-13.7)
G3 ^a	-223.9 (0.9)		-158.4 (0.0)		-24.6 (1.7)		-384.9 (-1.1)	-27.2 (-0.3)
Expt. ^b	-223.0		-158.4		-22.9		-386.0	-27.5

^a Theoretical errors in parentheses are given by (Expt. – Theor.). ^b Taken from ref. 29 and 40.

is not well described such that B88 is too repulsive as compared to HF theory and PW91 and PBE96 exchanges lead to artificial bonding interactions.¹⁸ A promising approach to reduce the errors of DFT is the inclusion of an empirical dispersion correction.^{91,92} Truhlar's M06 family of functionals¹⁵ also perform satisfactorily for non-bonded interactions, extending greatly the applicability of DFT methods.

The improvement of the XYG3 functional over B3LYP is examined on several sets of noncovalent interactions from Zhao *et al.*^{15,34} as shown in Fig. 6. The NCIE31/05 set is made of (1) 6 hydrogen bond complexes (HB6), (2) 7 charge-transfer complexes (CT7), (3) 6 dipole interaction complexes (DI6), (4) 7 weak interaction complexes (WI7) and (5) 5 π–π stacking complexes (PPS5). Our interaction energies are computed by using the 6-311+G(3df,2p) basis set without considering basis set superposition error corrections. The reference energies are taken from the Truhlar group database website.^{15,34} We see that the XYG3 functional performs satisfactorily well on all 5 subsets, including PPS(5).³⁹ It was shown before that B2PLYP was unable to describe the π–π interaction complexes properly.⁹¹ MAD for PPS5 was 2.68 kcal mol⁻¹ for B2PLYP, as opposed to MAD of 0.25 kcal mol⁻¹ for XYG3.³⁹ It was believed that this was because the PT portion (~27%) was not big enough to overcome the repulsion from the DFT part.⁹¹ But we suggest that it is due to the fact that the orbitals from the truncated DFT in B2PLYP stray too far away from the real KS orbitals.

The geometries used for B3LYP and XYG3 calculations shown in Fig. 6 are taken from the Truhlar group database website^{15,34} optimized with the QCISD method. In Fig. 7, we compare the XYG3 calculated potential energy curves for benzene dimers with those of Sherrill's,⁹³ extrapolated to the CCSD(T)/aug-cc-pVQZ* level of theory. Here "*" denotes that g functions on C atoms and f functions on H atoms have been removed. For a sandwich benzene dimer, the CCSD(T) minimum occurs at 3.9 Å, with an interaction energy of -1.7 kcal mol⁻¹. The XYG3 minimum is slightly shorter (3.8 Å) and the calculated interaction is slightly weaker (-1.2 kcal mol⁻¹). For a T-shaped benzene dimer, the CCSD(T) and XYG3 minima are 5.0 and 4.9 Å, respectively, the corresponding interaction energies are -2.6 and -2.9 kcal mol⁻¹, respectively. Fig. 7c shows the situation of a parallel-displaced benzene dimer at R₁ = 3.2 Å. The XYG3 curve is nearly on-top of the CCSD(T) curve. Both predict R₂ = 1.8 Å, and the interaction energies are -1.8 for CCSD(T) and -2.1 kcal mol⁻¹ for XYG3. The results displayed in Fig. 7 suggest that XYG3 is able to give reliable geometries and interaction energies for PPS systems.

4. Summary and outlook

Density functional theory (DFT) is now the leading first-principles method for electronic structure calculations in quantum chemistry. Various approximations to the exchange–correlation energy have been developed and tested in recent decades.^{5–20} In view of Perdew,⁹⁴ the hierarchy of density functional approximations can be pictured as “Jacob's ladder” rising from the “earth of Hartree” to the “heaven of chemical accuracy”. The first, second and third rungs of this ladder are LDA, GGA and *meta*-GGA. This fourth rung functional is the hybrid functional, such as B3LYP, which enjoys general popularity. The fifth rung of Jacob's ladder should utilize not only the occupied KS orbitals, but also the unoccupied KS orbitals. This final rung is expected to be close to the heaven of chemical accuracy for broad applications.^{95–97}

Here we propose a semiempirical fifth rung functional, XYG3,³⁹ which incorporates the information of unoccupied KS orbitals through Görling and Levy's coupling-constant perturbation expansion to the second order.⁴⁹ We demonstrate that XYG3 obviates the size dependence of B3LYP and is remarkably accurate not only for thermochemistry, but also for reaction barrier heights and nonbonded interactions.

It is anticipated that B3LYP as the unanimous No.1 choice has come to its final days, but B3LYP will remain a valid option for ‘every-day’ quantum chemistry problems.²⁸ The X1 method provides a systematic correction scheme, such that B3LYP data, already and continuously built-up in the literature, can be used with higher accuracy and thus higher reliability at no extra cost as compared to B3LYP. X1 significantly eliminates the notorious size dependent errors of B3LYP in prediction of heats of formation for larger molecules, and displays a significant improvement over B3LYP for bond dissociation energy and reaction barrier height predictions.

There are several limitations for both X1 and XYG3. (1) The X1 method sets up a neural-network to correct the B3LYP heats of formation at the level of B3LYP/6-311+G(3df,2p). Other versions, based on different functionals and basis sets, are in progress. (2) The price one pays for the introduction of MP2-like correlation energy in the doubly hybrid functionals is that one inherits the slow basis set convergence of dynamical correlation as that in wavefunction *ab initio* theory.^{20,26,91,98–100} The basis set dependence of XYG3 needs to be carefully evaluated. (3) The present version of X1 uses only atom types as descriptors and thus can be generally applied. Inclusion of new descriptors of bond and group types is expected to further improve the accuracy,

especially for isomerization energy predictions. (4) The present version of X1 is not applicable to charged species and non-covalent bondings. Instead, one may use X1 to correct the B3LYP errors for the neutrals and add B3LYP IPs (ionization potentials) or EAs (electron affinities) to get heats of formation for charged species. (5) Both X1 and XYG3 are trained and validated within the main group element chemistry, extension to the transition element chemistry is the next step to go. (6) Only single point energy calculations are available at the present time for both X1 and XYG3, which may fail if the B3LYP geometries degrade significantly. (7) The PT2 term scales formally as N^5 . We anticipate that linear scaling methods developed in the second-order Møller–Plesset perturbation theory^{101–104} can be readily used in our method for efficient calculations of large molecules.

Finally, we would like to emphasize that there are outliers for any approximate functionals and the quest for the divine functional is continuing. Grimme has designed an exchange reaction between the valence isoelectronic halides CX_4 and SiX_4 ($X = F$ and Cl), *i.e.*, $CF_4 + SiCl_4 \rightarrow CCl_4 + SiF_4$, and argued that a functional, capable of dealing with some ‘hard’ problems, but unable to deal with this ‘simple’ exchange problem, is a bad sign for lacking of ‘robustness’, and thus deserves no further attention by chemists.²⁰ We summarize our results in Table 8. B3LYP has large errors (up to $-20.3 \text{ kcal mol}^{-1}$) in the prediction of heats of formation of CX_4 and SiX_4 , but the errors tend to compensate with each other, to some extent, leading to a reduced error of $-10.9 \text{ kcal mol}^{-1}$ for the exchange reaction enthalpy. Indeed, this seemingly ‘simple’ reaction is not that simple as shown by the G2 calculations. Errors in G2 heats of formation accumulate, resulting in an error of $-13.7 \text{ kcal mol}^{-1}$ in the reaction enthalpy. In this respect, G3 is very satisfactory. We notice that XYG3 only reduces the B3LYP error for heats of reaction by 2 kcal mol^{-1} . Most of its error comes from its poor performance on the description SiF_4 . It is encouraging to see that X1 removes most B3LYP errors, giving heats of reaction in error by only $-1.5 \text{ kcal mol}^{-1}$. We note in passing that M06-2x leads to $-12.3 \text{ kcal mol}^{-1}$ in prediction of this exchange reaction. As errors are inevitable with an approximate functional, we anticipate neural-network corrections on top of XYG3 would provide an efficient way to achieve even higher accuracy with a broad applicability for chemistry.

Acknowledgements

This work is supported by NSFC (20525311, 20923004, 10774126, 20973138) and the Ministry of Science and Technology (2007CB815206).

Notes and references

- D. P. Chong, *Recent Advances in Density Functional Methods*, Pt. I and II, World Scientific, Singapore, 1997.
- V. Barone and A. Bencini, *Recent Advances in Density Functional Methods*, Pt. III, World Scientific, Singapore, 1999.
- W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley-VCH Verlag GmbH, Weinheim, 2001.
- P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864.
- J. C. Slater, *Quantum Theory of Molecules and Solids*, McGraw-Hill, New York, 1974, vol. 4.
- S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200.
- A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098.
- C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter*, 1988, **37**, 785.
- J. P. Perdew and Y. Wang, *Phys. Rev. B: Condens. Matter*, 1992, **45**, 13244.
- J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.
- Y. Zhang, A. A. Wu, X. Xu and Y. J. Yan, *Chem. Phys. Lett.*, 2006, **421**, 383.
- X. Xu and W. A. Goddard, III, *J. Chem. Phys.*, 2004, **121**, 4068.
- J. M. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 146401.
- T. Van Voorhis and G. E. Scuseria, *J. Chem. Phys.*, 1998, **109**, 400.
- Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215.
- A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372.
- A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648.
- X. Xu and W. A. Goddard III, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 2673.
- P. Mori-Sanchez, A. J. Cohen and W. T. Yang, *J. Chem. Phys.*, 2006, **124**, 091102.
- S. Grimme, *J. Chem. Phys.*, 2006, **124**, 034108.
- N. C. Handy and A. J. Cohen, *Mol. Phys.*, 2001, **99**, 403.
- X. Xu and William A. Goddard III, *J. Phys. Chem. A*, 2004, **108**, 8495.
- C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158.
- C. Adamo and V. Barone, *J. Chem. Phys.*, 1998, **108**, 664.
- A. D. Boese and J. M. L. Martin, *J. Chem. Phys.*, 2004, **121**, 3405.
- A. Karton, A. Tarnopolsky, J. -F. Lamère, G. C. Schatz and J. M. L. Martin, *J. Phys. Chem. A*, 2008, **112**, 12868.
- W. Kohn, A. D. Becke and R. G. Parr, *J. Phys. Chem.*, 1996, **100**, 12974.
- S. F. Sousa, P. A. Fernandes and M. J. Ramos, *J. Phys. Chem. A*, 2007, **111**, 10439.
- L. A. Curtiss, K. Raghavachari, P. C. Redfern and J. A. Pople, *J. Chem. Phys.*, 2000, **112**, 7374.
- M. D. Wodrich, C. Corminboeuf and P. v. R. Schleyer, *Org. Lett.*, 2006, **8**, 3631.
- C. E. Check and T. M. Gilbert, *J. Org. Chem.*, 2005, **70**, 9828.
- E. I. Izgorodina, M. L. Coote and L. Radom, *J. Phys. Chem. A*, 2005, **109**, 7558.
- P. R. Schreiner, A. A. Fokin, R. A. Pascal, Jr. and A. De Meijere, *Org. Lett.*, 2006, **8**, 3635.
- Y. Zhao, N. González-García and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 2012.
- Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 5656.
- J. M. Wu and X. Xu, *J. Chem. Phys.*, 2007, **127**, 214105.
- J. M. Wu and X. Xu, *J. Chem. Phys.*, 2008, **129**, 164103.
- J. M. Wu and X. Xu, *J. Comput. Chem.*, 2009, **30**, 1424.
- Y. Zhang, X. Xu and W. A. Goddard, III, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 4963.
- L. A. Curtiss, K. Raghavachari, G. W. Trucks and J. A. Pople, *J. Chem. Phys.*, 1991, **94**, 7221.
- W. Kohn and J. L. Sham, *Phys. Rev.*, 1965, **140**, A1133.
- Y. K. Zhang and W. T. Yang, *J. Chem. Phys.*, 1998, **109**, 2604.
- J. F. Dobson, K. McLennan, A. Rubio, J. Wang, T. Gould, H. M. Le and B. P. Dinte, *Aust. J. Chem.*, 2001, **54**, 513.
- S. Grimme, *Angew. Chem., Int. Ed.*, 2006, **45**, 4460.
- D. C. Langreth and J. P. Perdew, *Phys. Rev. B: Solid State*, 1977, **15**, 2884.
- O. Gunnarsson and B. Lundqvist, *Phys. Rev. B: Solid State*, 1976, **13**, 4274.
- S. Kurth and J. P. Perdew, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1999, **59**, 10461.
- J. P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982.
- A. Görling and M. Levy, *Phys. Rev. B: Condens. Matter*, 1993, **47**, 13105.
- B. Delley, *J. Phys. Chem. A*, 2006, **110**, 13632.
- J. Tirado-Rives and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2008, **4**, 297.
- P. Winget and T. Clark, *J. Comput. Chem.*, 2004, **25**, 725.

- 53 J. Cioslowski, G. Liu and P. Piskorz, *J. Phys. Chem. A*, 1998, **102**, 9890.
- 54 D. A. Long and J. B. Anderson, *Chem. Phys. Lett.*, 2005, **402**, 524.
- 55 R. A. Friesner, E. H. Knoll and Y. Cao, *J. Chem. Phys.*, 2006, **125**, 124107.
- 56 X. M. Duan, G. L. Song, Z. H. Li, X. J. Wang, G. H. Chen and K. N. Fan, *J. Chem. Phys.*, 2004, **121**, 7086.
- 57 Hu, X. Wang, L. Wong and G. H. Chen, *J. Chem. Phys.*, 2003, **119**, 11501.
- 58 J. M. L. Martin and G. de Oliveira, *J. Chem. Phys.*, 1999, **111**, 1843.
- 59 J. W. Ochterski, G. A. Petersson and J. A. Montgomery Jr., *J. Chem. Phys.*, 1996, **104**, 2598.
- 60 S. Kumar, *Neural Networks*, McGraw-Hill Companies Inc., 2006.
- 61 D. R. Hush and B. G. Horne, *IEEE Signal Process. Mag.*, 1993, **10**, 8.
- 62 X. Yao, *Proc. IEEE*, 1999, **87**, 1423.
- 63 C. J. Zhang, X. Xu and Q. E. Zhang, *Chem. Phys. Lett.*, 2002, **364**, 213.
- 64 M. W. Chase Jr., C. A. Davies, J. R. Downey, Jr., D. J. Frurip, R. A. McDonald and A. N. Syverud, *J. Phys. Chem. Ref. Data*, 1985, **14**, 1.
- 65 B. Ruscic, J. E. Boggs, A. Burcat, A. G. Csaszar, J. Demaison, R. Janoschek, J. M. L. Martin, M. L. Morton, M. J. Rossi, J. F. Stanton, P. G. Szalay, P. R. Westmoreland, F. Zabel and T. Berces, *J. Phys. Chem. Ref. Data*, 2005, **34**, 573.
- 66 P. C. Redfern, P. Zapol and L. A. Curtiss, *J. Phys. Chem. A*, 2000, **104**, 5850.
- 67 H. Y. Afeefy, J. F. Liebman and S. E. Stein, in *Neutral Thermochemical Data*, NIST Chemistry WebBook NIST Standard Reference Database Number 69, ed. P. J. Linstrom and W. G. Mallard, National Institute of Standards and Technology, Gaithersburg, MD, 20899, June 2005 (<http://webbook.nist.gov>).
- 68 S. J. Blanksby and G. B. Ellison, *Acc. Chem. Res.*, 2003, **36**, 255.
- 69 D. F. McMillen and D. M. Golden, *Annu. Rev. Phys. Chem.*, 1982, **33**, 493.
- 70 J. A. M. Simoes and J. L. Beauchamp, *Chem. Rev.*, 1990, **90**, 629.
- 71 L. A. Curtiss, P. C. Redfern, K. Raghavachari, V. Rassolov and J. A. Pople, *J. Chem. Phys.*, 1999, **110**, 4703.
- 72 S. Gronert, *Chem.–Eur. J.*, 2009, **15**, 5372.
- 73 J. Howell, J. D. Goddard and W. Tam, *Tetrahedron*, 2009, **65**, 4562.
- 74 M. D. Wodrich, C. S. Wannere, Y. Mo, P. D. Jarowski, K. N. Houk and P. v. R. Schleyer, *Chem.–Eur. J.*, 2007, **13**, 7731.
- 75 M. Lattalais, F. Pauzat, Y. Ellinger and C. Ceccarelli, *Astrophys. J.*, 2009, **696**, L133.
- 76 H. L. Woodcock, H. F. Schaefer III and P. R. Schreiner, *J. Phys. Chem. A*, 2002, **106**, 11923.
- 77 X. Q. You, F. N. Eglfopoulos and H. Wang, *Proc. Combust. Inst.*, 2009, **32**, 403.
- 78 G. Fu, X. Xu, X. Lu and H. L. Wan, *J. Phys. Chem. B*, 2005, **109**, 6416.
- 79 X. Xu, F. Faglioni and W. A. Goddard III, *J. Phys. Chem. A*, 2002, **106**, 7171.
- 80 X. Xu, S.-Y. Kang and T. Yamabe, *Chem.–Eur. J.*, 2002, **8**, 5351.
- 81 X. Xu, S.-Y. Kang and T. Yamabe, *Phys. Rev. Lett.*, 2002, **88**, 076106.
- 82 J.-M. Lehn, *Supramolecular Chemistry. Concepts and Perspective*, VCH, Weinheim, 1995.
- 83 N. Kannan and S. Vishveshwara, *Protein Eng., Des. Sel.*, 2000, **13**, 753.
- 84 H.-M. Zhang, Z.-X. Xie, B.-W. Mao and X. Xu, *Chem.–Eur. J.*, 2004, **10**, 1415.
- 85 G. R. Jenness and K. D. Jordan, *J. Phys. Chem. C*, 2009, **113**, 10242.
- 86 A. I. Kolesnikov, J.-M. Zanotti, C.-K. Loong, P. Thiyagarajan, A. P. Moravsky, R. O. Loutfy and C. J. Burnham, *Phys. Rev. Lett.*, 2004, **93**, 035503.
- 87 S. Grimme, *Angew. Chem., Int. Ed.*, 2008, **47**, 3430.
- 88 P. Hobza and J. Šponer, *Chem. Rev.*, 1999, **99**, 3247.
- 89 K. S. Kim, P. Tarakeshwar and J. Y. Lee, *Chem. Rev.*, 2000, **100**, 4145.
- 90 J. Cerny and P. Hobza, *Phys. Chem. Chem. Phys.*, 2007, **9**, 5291.
- 91 T. Schwabe and S. Grimme, *Phys. Chem. Chem. Phys.*, 2007, **9**, 3397.
- 92 Q. Wu and W. T. Yang, *J. Chem. Phys.*, 2002, **116**, 515.
- 93 M. O. Sinnokrot and C. D. Sherrill, *J. Phys. Chem. A*, 2004, **108**, 10200.
- 94 J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria and G. I. Csonka, *J. Chem. Phys.*, 2005, **123**, 062201.
- 95 A. Ruzsinszky, J. P. Perdew and G. I. Csonka, *J. Chem. Theory Comput.*, 2010, **6**, 127.
- 96 L. A. Constantin, J. M. Pitarke, J. F. Dobson, A. Garcia-Lekue and J. P. Perdew, *Phys. Rev. Lett.*, 2008, **100**, 036401.
- 97 J. Dobson, J. Wang and T. Gould, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2002, **66**, 081108.
- 98 A. Tarnopolsky, A. Karton, R. Sertchook, D. Vuzman and J. M. L. Martin, *J. Phys. Chem. A*, 2008, **112**, 3.
- 99 J. C. Sancho-García and A. P. Pérez-Jiménez, *J. Chem. Phys.*, 2009, **131**, 084108.
- 100 D. C. Graham, A. S. Menon, L. Goerigk, S. Grimme and L. Radom, *J. Phys. Chem. A*, 2009, **113**, 9861.
- 101 M. Kobayashi, Y. Imamura and H. Nakai, *J. Chem. Phys.*, 2007, **127**, 074103.
- 102 M. Häser and J. Almlöf, *J. Chem. Phys.*, 1992, **96**, 489.
- 103 G. Rauhut and P. Pulay, *Chem. Phys. Lett.*, 1996, **248**, 223.
- 104 P. Y. Ayala and G. E. Scuseria, *J. Chem. Phys.*, 1999, **110**, 3660.