Theor Chem Acc (2004) 112: 1–6
DOI 10.1007/s00214-003-0555-y

Theoretical
Chemistry Accounts

Accurate energies calculated by empirical corrections to the local spin density approximation

René Fournier, Luba Boroukhovskaia

Department of Chemistry, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada

Received: 31 March 2003 / Accepted: 28 July 2003 / Published online: 8 March 2004

Abstract. We tested the efficacy of three empirical correction schemes on atomization energies calculated by the following theories: Kohn–Sham density functional theory (KS-DFT) with local spin density approximation (LSDA), two KS-DFT gradient-corrected methods, one hybrid Hartree–Fock/KS-DFT method similar to B3LYP, and the ab initio extrapolation procedures G1 and G2. Empirical corrections improved the LSDA results greatly, while the other theories were improved slightly or not at all. The best procedure for correcting LSDA atomization energies brings the mean absolute deviation from experiment from 38.3 to 4.0 kcal/mol on a subset of 44 molecules in the G1 dataset that were not used in deriving the empirical parameters. This corrected LSDA is interesting for three reasons: it could be a useful computational tool in some cases, it implies that the LSDA itself gives accurate energies for reactions where atomic coordinations stay unchanged, and it gives insight into the search of better functionals.

Keywords: Local spin density – Thermochemistry – Empirical corrections – Functionals

1 Introduction

The local spin density approximation (LSDA) [1, 2, 3] is the starting point for almost all exchange–correlation functionals used in Kohn–Sham (KS) density functional theory (DFT) [4]. The LSDA is notorious for its poor accuracy on bond energies, but it is remarkably good for molecular properties that depend only on small regions of the potential–energy surface near minima such as equilibrium geometries [5] and harmonic vibrational frequencies [6]. This suggests that the LSDA gives a rather good description of molecules, but a poor description of atoms, and that errors in LSDA atomization energies could largely be removed by a simple shift of reference atomic energies [7]. We tested this idea, and two other types of empirical corrections, for the LSDA and five other methods – two gradient-corrected DFTs, a hybrid Hartree–Fock/DFT, and two ab initio extrapolation schemes.

The so-called G1 dataset [8] includes 55 small molecules of main group elements ($Z < 18$) with accurately known atomization energies. It has become a standard reference for assessing different theories for thermochemistry. The atomization energy of a molecule, which we denote by D_0, is a fundamental quantity for thermochemistry. A possible estimation of the accuracy of a theory is given by the mean absolute deviation (MAD) of its predicted D_0 from the highly accurate experimental values of the G1 dataset. The MAD for LSDA, for instance, is 36 kcal/mol—unacceptably large for typical applications.

The failures of the LSDA in thermochemistry were recognized early on, and much effort went into devising improved exchange–correlation functionals giving better energies. Becke's 1992 papers [9, 10] were very important, as they showed that gradient-corrected functionals can give MADs from the G1 dataset experimental atomization energies of 4–6 kcal/mol [10]. Such generalized gradient approximation (GGA) functionals are now better understood, and they have been refined and simplified, see, for example, Ref. [11]. Shortly after, Becke obtained an even more accurate functional, with a MAD of only 2.4 kcal/mol, by mixing exact Hartree–Fock exchange with a DFT exchange–correlation functional [12]. Some of the newer functionals use the orbital kinetic energy density and can deliver high accuracy even though they contain no Hartree–Fock exchange contribution and few adjustable parameters. For instance, one functional by Perdew et al. [13] has only two adjustable parameters and gives a MAD of 3.1 kcal/mol on a set of 20 atomization energies. Other DFT methods achieve higher accuracy but at the cost of increased complexity and the introduction of many adjustable parameters

---

Correspondence to: R. Fournier
e-mail: renef@yorku.ca

[14]. Some quantum chemical methods extrapolate energies from sequences of ab initio calculations in which the basis set and description of electron correlation get progressively more complete. Such methods yield extremely accurate results, with MADs ranging from 1 to 2 kcal/mol [8, 15, 16, 17]. However, they are significantly more costly in computing time, and the theory underlying them does not perform as well for transition-metal elements as it does for the light elements that are predominant in test datasets. There is continuing interest in trying to improve the accuracy of DFT methods because they are computationally more efficient than conventional ab initio methods and their accuracy for systems with heavier elements is encouraging.

A quick look at D_0 from experiment and from LSDA calculations (see Ref. [9] or Table 1) shows simple trends, for instance, the LSDA almost always overestimates atomization energies. Although the LSDA values of D_0 are very poor in absolute terms, they look rather good in a relative sense and there could be simple corrections that bring them in much better agreement with experiment. The purpose of this article is to discuss the relative accuracy of different theories for predicting D_0 and to assess how different empirical corrections may help improve theoretical methods. From our results (Section 3) it appears that the LSDA is a much better theory than is generally acknowledged, and that the LSDA may even be a reasonable choice for certain kinds of thermochemical calculations.

Table 1. Local spin density approximation (*LSDA*) atomization energies D_0 (kcal/mol) without and with corrections of Eqs. (8) and (9)

Molecule	Expt	LSDA	LSDA-c2	LSDA-c3
CH	79.9	88.6	50.7	68.8
CH_2 triplet	179.6	202.4	162.7	173.2
CH_2 singlet	170.6	188.9	149.2	159.7
CH_3	289.2	321.7	280.1	285.0
CH_4	392.5	435.5	392.1	392.1
NH	79.0	91.0	54.6	69.2
NH_2	170.0	196.5	158.2	164.6
OH	101.3	119.1	86.5	95.5
Li_2	24.0	23.0	21.1	21.1
LiF	137.6	154.2	133.7	133.7
C_2H_2	388.9	443.6	367.9	389.0
C_2H_4	531.9	601.7	522.3	531.9
CN	176.6	217.4	146.9	179.4
HCN	301.8	351.1	278.7	303.8
CO	256.2	296.1	229.4	256.4
HCO	270.3	325.1	256.6	276.1
H_2CO	357.2	417.6	347.2	361.0
CH_3OH	480.8	555.4	481.3	481.3
N_2	225.1	264.6	195.5	224.7
N_2H_4	405.4	483.8	407.3	407.3
NO	150.1	196.3	131.1	154.6
O_2	118.0	172.5	111.1	129.1
H_2O_2	252.3	318.9	253.8	253.8
F_2	36.9	76.7	37.6	37.6
CO_2	381.9	466.2	368.8	397.3
SiH_2 singlet	144.4	159.2	136.5	142.1
SiH_2 triplet	123.4	139.5	116.8	122.4
SiH_3	214.0	233.6	209.0	211.6
SiH_4	302.8	327.9	301.5	301.5
PH_2	144.7	165.7	139.7	149.1
Si_2	74.0	92.8	54.8	73.8
P_2	116.1	143.1	98.4	117.3
S_2	100.7	134.6	93.8	105.8
Cl_2	57.2	82.8	58.3	58.3
NaCl	97.5	102.8	88.8	88.8
SiO	190.5	222.7	173.0	191.5
CS	169.5	200.6	144.2	168.2
SO	123.5	166.4	115.3	130.3
ClO	63.3	104.0	61.0	70.0
ClF	60.3	94.3	62.5	62.5
CH_3Cl	371.0	424.7	370.9	370.9
CH_3SH	445.1	508.5	444.7	444.7
HOCl	156.3	203.2	158.4	158.4
SO_2	254.0	332.8	251.0	269.0

Kristyán et al. [18, 19] recently proposed empirical corrections to Hartree–Fock energies that greatly improve calculated atomization energies. The kind of empirical corrections that we consider here are different from theirs in many ways. Our starting point is the LSDA instead of Hartree–Fock, but we also examined the effect of corrections to higher levels of theory. Our corrections involve only one parameter per element and they depend only on molecular geometry, not on atomic charges [18, 19] or other information derived from orbitals or electron density. We did not restrict our tests to closed-shell molecules: we looked at open-shell molecules and find that they give the largest errors but also some valuable insight. Finally, our corrections (Eqs. 7, 8, 9) are designed so that they become exactly zero for a certain class of reactions. Therefore, the performance of corrected theories on atomization energies gives a sense for the expected accuracy of uncorrected theories for energies of reactions where the corrections are exactly zero.

2 Empirical corrections

We compared six theories with experiment on the basis of the atomization energy D_0 of molecules in the G1 dataset [8]: the LSDA [3]; the Bx exchange-only functional of Becke [20]; the same functional with the gradient corrections of Perdew and Wang to correlation [21] added, which we denote BPW; the three-parameter hybrid functional B3PW of Becke [12], which calculates the exchange energy as a sum of Hartree–Fock and DFT contributions and is very similar to the popular B3LYP functional; and the ab initio based extrapolation procedures G1 and G2. There are many more computational methods, to be sure, but the ones considered here are a representative sample of practical methods used for thermochemical calculations. We took atomization energies previously published [8, 9, 12, 15]. The DFT calculations were done with the basis-set-free program NUMOL [9, 10], and the G1 and G2 results are extrapolations based on several calculations [8, 15]. In all cases, the results correspond to a nearly complete basis set treatment and have a very small numerical uncertainty. In each case we compared theoretical values (y_i) with experimental ones (x_i) by calculating deviations $d_i = y_i - x_i$ and relative deviations $\delta_i = (y_i - x_i)/x_i$. We calculated the following measures of accuracy:

$$\text{Bias} = \text{mean}(d_i) \tag{1}$$

$$\mathrm{MAD} = \mathrm{mean}(|d_i|) \ , \tag{2}$$

$$\mathrm{RMSD} = [\mathrm{mean}(d_i^2)]^{1/2} \ , \tag{3}$$

$$\mathrm{rMAD} = 100 \times \mathrm{mean}(|\delta_i|) \ , \tag{4}$$

$$\mathrm{rRMSD} = 100 \times [\mathrm{mean}(\delta_i^2)]^{1/2} \ , \tag{5}$$

$$\beta = (\mathrm{RMSD} - |\mathrm{bias}|)/\mathrm{MAD} \ . \tag{6}$$

The MAD is what is most often quoted in the literature. The root-mean-square deviation (RMSD) is more sensitive to the presence of a few large deviations. The rMAD and rRMSD, expressed as percentages, are similar to MAD and RMSD, respectively, but they magnify the effect of errors on molecules with a smaller D_0. We must always have $|\mathrm{bias}| \leq \mathrm{MAD} \leq \mathrm{RMSD}$, and the limit case $|\mathrm{bias}| = \mathrm{MAD} = \mathrm{RMSD}(\beta = 0)$ is attained when the error is constant, which allows a trivial correction. Generally, smaller values of β indicate errors that are more systematic and easier to correct. We considered three kinds of empirical corrections:

$$D_0(\mathrm{c1}) = fD_0, \\ f = \sum_i n_i \Delta_1(X_i) / \sum_i n_i \ , \tag{7}$$

$$D_0(\mathrm{c2}) = D_0 - \sum_i \Delta_2(X_i) \ , \tag{8}$$

$$D_0(\mathrm{c3}) = D_0 - \sum_i \Delta_3(X_i) \times n_i^{1/2} \ . \tag{9}$$

The sums in Eqs. (7), (8) and (9) run over all atoms i in the molecule, and X_i identifies the element. The Δ's are the adjustable parameters of each model, with one parameter per element. The first correction is just a scaling by a factor f, which is a weighted average of atomic factors $\Delta_1(X_i)$ and where the atomic coordinations n_i are used as weights. In the second model, the energy of each atom is shifted by a fixed amount that depends only on the element. In the third model, we assume that the correction associated with an atom depends on its environment in the molecule, and that it goes as the square root of the coordination n_i of that atom. For the molecules considered here it is easy to assign n_i values. For arbitrary geometries one would have to use a well-defined criterion for calculating n_i. For example, one could consider all pairs (i, j) of atoms, and increment n_i and n_j by 1 whenever the distance R_{ij} between atoms i and j is less than $c(R_i^{\mathrm{cov}} + R_j^{\mathrm{cov}})$, the sum of their covalent radii multiplied by some constant c that is slightly larger than 1. In order to map out a potential surface, one would need n_i's that are continuous and differentiable with respect to nuclear coordinates. These practical considerations about the definition of n_i are not very important here because we look at energy differences between molecules at their equilibrium geometries for which the n_i's are unambiguous.

We did not fit the empirical parameters Δ of Eqs. (7), (8) and (9) to the G1 dataset or other datasets. Instead, we fixed the 11 parameters so as to reproduce the experimental atomization energies of eleven closed-shell molecules with formula X_nH_m, except for Na, where we took Na_2 because D_0(NaH) is not known accurately. The 11 molecules used to fix parameters are H_2, LiH, C_2H_6, NH_3, H_2O, HF, Na_2, Si_2H_6, H_2S, PH_3, and HCl. The D_0 for H_2 is not reported in Ref. [15], so we assumed that G1 and G2 agree perfectly with experiment for $D_0(H_2)$ and calculated the remaining empirical parameters as for the other methods.[1]

Table 2. Energy shift parameters $\Delta_3(E;X,\mathrm{LSDA})$ of Eq. (9), in (kcal/mol). The *numbers* in *parentheses* are from the two-parameter fit Eq. (10)

Element $X/\Delta_3(X,\mathrm{LSDA})$ parameter					
H	Li	C	N	O	F
1.85	0.95	18.00	19.95	21.71	19.55
(3.89)	(1.01)	(16.17)	(20.10)	(20.72)	(18.45)
	Na	Si	P	S	Cl
	1.70	9.50	12.90	14.42	12.25
	(0.81)	(11.22)	(14.22)	(14.83)	(13.42)

Many empirical parameters enter Eqs. (7), (8) and (9), one per element represented in the dataset, and ultimately one for each element in the periodic table: However, these parameters have a simple meaning and require no fit so they can be obtained easily for any theory and computational model (software, basis set, etc.). Note also that these parameters follow simple trends across the periodic table, so they can themselves be fitted to an empirical formula having fewer parameters. For example, the 11 energy shift parameters of Eq. (9) for the LSDA (Table 2) can be fitted rather well by a formula with two adjustable parameters:

$$\Delta_3 \approx aV^b \left[\frac{(n_\alpha^{4/3} + n_\beta^{4/3})}{(4\pi/3)^{1/3} R} \right] \ . \tag{10}$$

In Eq. (10), n_α and n_β are the spin up and spin down numbers of valence electrons for the element in its ground-state configuration, $V = \min\{(n_\alpha + n_\beta), 8 - (n_\alpha + n_\beta)\}$ is the valence of the atom, R is the covalent atomic radius, and $a = 0.00604$ au and $b = 0.455$ are fitting parameters. The MAD for the 44 test molecules is 5.8 kcal/mol in this two-parameter model, not much more than with the underlying 11-parameter model (4.0 kcal/mol).

The term in brackets in Eq. (10) is proportional to the exchange energy in an ultrasimplified model of an atom where only valence electrons are counted and the spin densities are assumed constant inside a sphere of radius R. One expects a relation between the magnitude of Δ_3 and an atom's model exchange energy. The proportionality factor, aV^b, represents the lack of error cancellation in energy between the free atom and the atom in the molecule, expressed as a fraction of the atom's model exchange energy. The dependence on V is not too surprising: V represents the number of electrons that are strongly affected by formation of chemical bonds.

---

[1] We get very similar results as reported here when we use CH_4 and C_2H_6, instead of H_2 and C_2H_6, for obtaining the empirical parameters of the H and C atoms for G1 and G2 theories.

Values of Δ_3 calculated by Eq. (10) are shown in parentheses in Table 2.

3 Results and discussion

We compare theory with experiment for the uncorrected theories (LSDA, Bx, BPW, B3PW, G1, G2) and for each theory empirically corrected using Eqs. (7), (8) and (9) in Table 3. In order to make a fair comparison, we eliminated the 11 molecules used to derive empirical parameters (and also BeH) from the G1 dataset. This leaves us with a test set of 44 molecules. The bottom part of Table 3 repeats the best empirically corrected results for each theory. First we look at uncorrected theories at the top of Table 3. The different measures of deviation from experiment, including rMAD and rRMSD, give the same ranking of theories. LSDA is by far the worst, with errors of the order of 40 kcal/mol, or 25%. Gradient-corrected DFT methods are much better, with errors of roughly 7 kcal/mol and 8% for BPW and 4 kcal/mol and 4% for Bx. Mixing exact exchange with DFT exchange–correlation (B3PW) further reduces the error to 2.5 kcal/mol or 4%. The accuracy of Bx and B3PW relative to BPW shows how important the treatment of exchange is for atomization energies. The G1 and G2 results are the most accurate, with errors of roughly 2 kcal/mol and 2% for G1 and 1.5 kcal/mol and 1.5% for G2. The errors are much more systematic in the LSDA than in other theories, as the small value of β indicates. The ratios (rMAD/MAD) and (rRMSD/RMSD) are smaller in the LSDA (0.59 and 0.66, respectively, compared with 0.8 and 1.1 on average for other theories), which also shows that the LSDA can be accurate in a relative sense. In contrast, the three theories with the lowest MAD have the largest β.

Table 3. Different measures of accuracy (Eqs.1, 2, 3, 6) of theories compared with 44 experimental atomization energies (kcal/mol or percent)

Theory	Bias	MAD	RMSD	β
		Uncorrected		
LSDA	38.3	38.3	43.6	0.1
Bx	2.3	4.0	4.7	0.6
BPW	4.4	6.2	7.9	0.6
B3PW	0.2	2.4	3.0	1.1
G1	−0.5	1.6	2.1	1.0
G2	−0.1	1.2	1.6	1.2
		Corrected by Eq. (7)		
LSDA	1.6	6.0	8.4	1.1
Bx	4.6	5.2	6.5	0.4
BPW	3.9	4.4	5.4	0.4
B3PW	0.0	2.4	3.1	1.3
G1	1.0	1.8	2.4	0.8
G2	−0.5	1.3	1.8	1.0
		Corrected by Eq. (8)		
LSDA	−11.2	11.6	15.0	0.3
Bx	9.2	9.6	11.3	0.2
BPW	2.1	3.2	4.2	0.7
B3PW	−1.2	2.7	3.7	0.9
G1	1.9	2.3	2.9	0.4
G2	−0.5	1.3	1.7	0.9
		Corrected by Eq. (9)		
LSDA	0.4	4.0	5.6	1.3
Bx	6.2	6.7	7.7	0.2
BPW	3.3	3.8	5.0	0.5
B3PW	−0.5	2.3	3.0	1.1
G1	1.3	1.8	2.2	0.5
G2	−0.5	1.2	1.6	0.9
		Best corrected theories		
LSDA−c3	0.37	3.96	5.63	1.33
Bx	2.28	3.97	4.70	0.61
BPW-c2	2.10	3.16	4.21	0.67
B3PW-c3	−0.48	2.34	2.97	1.06
G1	−0.49	1.55	2.06	1.01
G2-c3	−0.46	1.22	1.60	0.93

The correction of Eq. (7) improves the LSDA a lot (the MAD drops to 6.0 kcal/mol) but it does not improve the accuracy of other theories much. Curiously, using a single scaling parameter, fixed in a way that eliminates the bias, decreases slightly the MAD in Bx theory (from 4.0 to 3.8 kcal/mol). The correction of Eq. (8) also improves the LSDA, but not as much; the MAD decreases to 11.6 kcal/mol. Errors on atomization energies can always be viewed as due to an incomplete cancellation of errors in the total energies of the atoms and the molecule. Since the LSDA gives accurate molecular equilibrium structures and vibrational frequencies, we thought that the LSDA error on D_0 is mainly due to a bad description of ground-state atoms. This is what motivated our choice of Eq. (8), and the large decrease in MAD with Eq. (8) (from 38 to 12 kcal/mol) seems to support this viewpoint. However, the corrections of Eqs. (7) and (9) give even better results than that, so LSDA errors on D_0 cannot simply be ascribed to errors on atomic energies. Surprisingly, Eq. (8) also cuts the error in BPW by a factor of almost 2. Clearly, if a chemical reaction has no atom among products and reactants, the correction of Eq. (8) to the energy of reaction is zero. Mean errors on energies of reaction are not equal to mean errors on atomization energies, but they are surely strongly correlated on account of Hess' law. Therefore, the good accuracy of BPW-c2 for D_0 suggests that BPW, and GGA methods generally, give better energies of reaction than the MAD over the G1 dataset indicates, provided there is no atom among the reactants or products.

The correction of Eq. (9) dramatically improves the LSDA! The bias is eliminated, and the MAD and RMSD on absolute and relative deviations decrease by factors of 6–10. By any measure, the LSDA corrected with Eq. (9) is nearly as good as uncorrected Bx and BPW. More remarkable still is that the LSDA corrected with Eq. (9) is only slightly worse (4.0 versus 3.2 kcal/mol) than the best empirically corrected GGA. We think it is close enough in accuracy to other DFT methods to deserve further studies, especially in systems with heavier atoms or metallic character where inclusion of some Hartree–Fock exchange, as is done in B3PW, could be detrimental.

The bottom part of Table 3 shows the best empirical version of each theory. The ranking, in decreasing order of deviation from experiment, is as for uncorrected theories except that BPW-c2 gets better than Bx:

LSDA-c3 = Bx > BPW-c2 > B3PW-c3 > G1 > G2-c3. This was expected: empirical corrections are not a substitute for a better treatment of exchange and correlation; however, empirical corrections bring the accuracy of various theories much closer to one another. The ratio of the largest to the smallest MAD is 31 among uncorrected theories, and it is 3.2 among their best empirical versions. The decrease in the error on D_0 on going to higher levels of theory is very big, but only a small fraction of it cannot be recovered by simple empirical corrections. If one is willing to use empirical corrections, then computational cost and reliability for a certain category of molecules or reactions become deciding factors in the choice of a computational method, and then the LSDA-c3 and BPW-c2 models look like interesting alternatives.

Empirical corrections are not satisfying from a theoretical viewpoint, but the success of Eq. (9) with the LSDA, in particular, has interesting implications. Firstly, it shows the kind of situations where one can expect the uncorrected LSDA to work well. Define atom "type" as the combination of an element and coordination: when the number of atoms of each type stays unchanged during a chemical reaction, corrections Eqs. (7) and (9) (and also Eq. 8) become identically zero[2]. So the calculated LSDA energy change for such a reaction should be roughly as accurate as LSDA-c3 is for atomization energies (4 kcal/mol). Secondly, it points to the possibility of improving the LSDA in different ways than the GGA approach. Nonzero density gradients are not the only aspect by which atoms and molecules differ fundamentally from the homogeneous electron gas (the basic model of the LSDA): their finite size is another. The form of Eqs. (9) and (10) suggests that a big part of the error in the LSDA comes from a lack of error cancellation in exchange energy contributions coming from regions of space that are inside versus outside effective surfaces bounding atoms and molecules. Indeed, Becke stressed the importance for thermochemistry of low-electron-density regions and devised gradient corrections that give correct asymptotic behavior in the exchange–correlation energy density [9]. It may be that simple functionals without density gradient, but with corrections that depend on some other variable instead, could also give accurate thermochemistry. One possibility might be a surface energy correction [22]; however, defining "surface energy" in atoms and molecules is not simple.[3] Our Eq. (9) seems to achieve roughly the same result that a functional with correct asymptotic form, like Bx, produces in a much more rigorous way. Thirdly, Eq. (9) could just be used as a practical tool for estimating reaction energies when other methods are too costly or impractical. If LSDA-c3 is to be used in that way, it is worth taking a closer look at errors.

Experimental and corrected LSDA atomization energies for the 44 test molecules are shown in Table 1. LSDA-c3 is extremely accurate for many of the closed-shell molecules but it still has deficiencies: it underestimates D_0 of small radicals with fewer than ten electrons (CH, CH_2, etc.); it overestimates D_0 in molecules containing singly coordinated oxygen or sulfur atoms. We can achieve much better accuracy at the cost of introducing more ad hoc corrections and three more parameters: (1) corrections for each singly coordinated O atom (−5.55 kcal/mol) and S atom (−2.55 kcal/mol) which make O_2 and S_2 exact; (2) a correction equal to $4.2\,(10 - N_e)$ for small radicals which makes the CH_3 radical exact. We call this last model LSDA-c4. With these additional corrections, the MAD measured against a reduced test set of 41 molecules (excluding CH_3, O_2 and S_2) becomes 2.0 kcal/mol. We do not recommend LSDA-c4 as a general method, but it might be used as an empirical method that needs testing before applications can be made to specific problems. The LSDA-c4 model has too many arbitrary corrections and parameters to be a good predictive tool. It would not be reliable for systems with ionic bonds (e.g., see LSDA-c3 errors on NaCl and LiF in Table 1), delocalized π bonds, metal–ligand bonds, etc., which are absent, or few, in the G1 dataset. Nevertheless, LSDA-c4 is an interesting model. Note that correction 1 is zero for a process where the number of atoms of each type stays unchanged (as are corrections by Eqs. 7, 8, 9), and correction 2 applies only to very small species for which high levels of theory are feasible and DFT is typically not needed. So one could expect the LSDA to give an accuracy even better than 4.0 kcal/mol for energies of reactions involving no light radicals and where the number of atoms of each type stays unchanged. The LSDA-c4 model shows that it is quite possible to calculate energy differences to high accuracy and at a small computational cost when empirical corrections are used and when applications are limited to systems with few elements and bonding mechanisms. A modest level of theory combined with simple corrections can be adequate for that purpose, as was also demonstrated by Kristyán et al. [18, 19].

4 Summary and conclusions

The LSDA gives errors on atomization energies that are large but systematic, as shown by the small β (Eq. 6, and Table 3). Corrections calculated by Eq. (9) bring the LSDA into much better agreement with experiment. These corrections are never zero for atomization energies, but they are often small or zero in typical applications of interest to chemists–energies of reactions. From that perspective, the LSDA appears much better than is generally stated in the literature. It may well give mean errors of 4 kcal/mol or even less on energies of reactions where atomic coordinations stay unchanged. This kind of accuracy is expected for a wider class of

---

[2] For example, $CH_2Cl_2 + H_2 \rightarrow CH_4 + Cl_2$. There are also reactions where the coordination of some atoms changes, but is compensated by an opposite change of coordination on other atoms of the same element, for example, $C_2H_4Cl_2 + C_2Cl_4 \rightarrow C_2H_4 + C_2Cl_6$.

[3] For instance, one could define the surface energy of an atom or molecule as the energy required to pull it out of a jellium of density ρ ("embedding energy"); but then, there is considerable leeway in how to choose ρ and no obvious criterion for fixing its value.

reactions when using empirical corrections as in LSDA-c3 or LSDA-c4 models. The BPW-c2 model (BPW theory) appears a very good practical tool: it simply requires shifts of reference atomic energies and gives a MAD of 3.2 kcal/mol.

The LSDA-c3 model in its two-parameter version is rather good (MAD of 5.8 kcal/mol). Its defining equations (Eqs. 9, 10) have an interesting mathematical form. They produce a sum of atom-in-molecule corrections, with each correction roughly proportional to $\sqrt{n_i V_i}$—the geometric average of the number of neighbours around atom i ($n_i^{1/2}$ in Eq. 9) and the number of valence electrons in atom i (V_i^b, $b = 0.455 \approx \frac{1}{2}$), which is akin to an effective number of chemical bonds. It is not clear exactly how this kind of result could arise naturally in a first-principles correction to the LSDA, but the simplicity and success of Eqs. (9) and (10) hint at the possible existence of simple and gradientless, yet accurate, functionals.

Acknowledgements. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

References

1. Slater JC (1972) Adv Quantum Chem 6: 1
2. Gunnarsson O, Lundqvist BI (1976) Phys Rev B 13: 4274
3. Vosko SH, Wilk L, Nusair M (1980) Can J Phys 58: 1200
4. Parr RG, Yang W (1989) Density-functional theory of atoms and molecules. Oxford University Press, New York
5. Dickson RM, Becke AD (1993) J Chem Phys 99: 3898
6. (a) Andzelm J, Wimmer E (1992) J Chem Phys 96: 1280; (b) Johnson BG, Gill PMW, Pople JA (1993) J Chem Phys 98: 5612; (c) Bérces A, Ziegler T (1993) J Chem Phys 98: 4793; (d) Fournier R, Pápai I (1995) In: Chong DP (ed) Recent advances in density functional methods, part I. World Scientific, Singapore pp 219–285; (e) Scott AP, Radom L (1996) J Phys Chem 100: 16502
7. Fournier R, Sinnott SB, DePristo AE (1992) J Chem Phys 97: 4149
8. (a) Pople JA, Head-Gordon M, Fox DJ, Raghavachari K, Curtiss LA (1989) J Chem Phys 90: 5622; (b) Curtiss LA, Jones C, Trucks GW, Raghavachari K, Pople JA (1990) J Chem Phys 93: 2537
9. Becke AD (1992) J Chem Phys 96: 2155
10. Becke AD (1992) J Chem Phys 97: 9173
11. (a) Perdew JP, Burke K, Ernzerhof M (1996) Phys Rev Lett 77: 3865; (b) Erratum (1997) Phys Rev Lett 78: 1396
12. Becke AD (1993) J Chem Phys 98: 5648
13. (a) Perdew JP, Kurth S, Zupan A, Blaha P (1999) Phys Rev Lett 82: 2544; (b) Erratum (1999) Phys Rev Lett 82: 5179
14. Schmider HL, Becke AD (1998) J Chem Phys 108: 9624
15. Curtiss LA, Raghavachari K, Trucks GW, Pople JA (1991) J Chem Phys 94: 7221
16. Curtiss LA, Raghavachari K, Redfern PC, Pople JA (2000) J Chem Phys 112: 7374
17. Benassi R (2001) Theor Chem Acc 106: 259
18. Kristyán S, Ruzsinszky A, Csonka GI (2001) Theor Chem Acc 106: 319
19. Kristyán S, Ruzsinszky A, Csonka GI (2001) Theor Chem Acc 106: 404
20. Becke AD (1988) Phys Rev A 38: 3098
21. (a) Perdew JP, Wang Y (1992) Phys Rev B 45: 13244; (b) Perdew JP, Chevary JA, Vosko SH, Jackson KA, Pederson MR, Singh DJ, Fiolhais C (1992) Phys Rev B 46: 6671
22. Almeida LM, Perdew JP, Fiolhais C (2002) 66: 075115