Simple Energy Corrections for Precise Atomization Energies of CHON Molecules

Pablo F. Salazar[†] and Jorge M. Seminario*,[†,‡]

Department of Chemical Engineering and Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas 77843

Received: June 25, 2007; In Final Form: August 5, 2007

We test a few ways to improve atomization energies of CHON molecules and found that the best way to do it is simply by correcting the atom energies of the participating atoms. Extraordinary improvement on the average errors is obtained. For the HF/6-31G** level of theory an average error of 271.2 kcal/mol on 115 molecules is improved to 6.7 kcal/mol simply by correcting the atomic energy of the four chon atoms. The corrections to density functional methods allow us to reach a chemical accuracy of 2 kcal/mol.

I. Introduction

Fitting parameters are commonly used to create new functionals; however, an alternative approach to simply correct the calculated energies, without attempting to redesign or modify the functionals is less common but very important to quickly calculate large systems. In the former case, the parameters actually modify the energy functionals from either theoretical or experimental feedback, whereas in the latter case, the parameters correct the energies found using the original procedures. As in previous work,[1] we follow the latter, noninvasive, methodology for our energy corrections; this approach consists of choosing contributions to the total energy thought to be the cause of errors; then, coefficients or weights, which we call correctors, are assigned to them. These correctors are found by minimizing the deviations with respect to precise theoretical or experimental values.

The nature of errors with one level of theory is usually different from the errors with other levels of theory, but they could be separately estimated. Also, errors due to the finite nature of the basis sets can be decomposed in contributions with respect to the angular moment of the basis functions. In most of the cases, the contributions are additive.[2,3]

Similar work has been already performed using correctors, for instance, Duan[4] used this methodology to correct for the electron correlation energy missing in the HF method by adding weighted descriptors to the HF energy, the weighting values were found by fitting to the experimental values of heat of formation energy using a linear regression; instead, we choose to fit to the dissociation energy by using a Newton minimization scheme. Also, Cioslowski[5] developed a bond density functional scheme, in which he introduced and calibrated additional parameters for the prediction of enthalpy of formation of large systems. Earlier work[6−8] toward chemical accuracy has been crucial to build present sophisticated correctors. More sophisticated minimization methods have also been tried; Wang[9] used a neural network-based scheme to find the weighting coefficients.

Recently, we introduced *a posteriori* multiplicity-based corrections to ab initio energies to reproduce experimental atomization energies.[1] This simple approach, compared to

the alternative ones to improve density functionals and standard correlated methods, also required fewer computational resources than higher levels of theory did. In this work, we tested our approach with 115 molecules of the G3/99 set.[10−12] The accuracy of the experimental energies of the test set was less than 1 kcal/mol, which is important for the development and validation of new methods. We limit our study to molecules composed of carbon, hydrogen, nitrogen, and oxygen, also known as CHON molecules.

II. Methodology

Several ab initio methods have been used in this work, aiming to evaluate the energy correctors through a broad span of levels of accuracy, from the inexpensive HF, going through the post-HF (MP2), through the DFT (B3PW91), and to the highly accurate G3x method. The Hartree−Fock method is the first approximation and the least expensive of the ab initio methods; the repulsion between electrons is taken into account as an average effect but not the specific repulsion interaction between electrons, which is known as the correlation effect. However, this approximation is still useful for first-order predictions of many systems. The correlation between electrons is considered by the Møller−Plesset MP2 theory,[13] which adds a second-order energy correction to the HF electronic energy. The B3PW91 uses the Becke-3 term hybrid exchange functional and the generalized-gradient approximation (GGA) Perdew−Wang 91 correlation functional.[14−16]

The G3 theory is a composite electronic structure method that is based on a sequence of *ab initio* molecular-orbital calculations. It approximates to a QCISD(T) high level of calculation at the MP2/6-31G* geometry and uses scaled HF/6-31G* for zero-point energies. The G3X theory[17] is an extension of the G3 theory that uses B3LYP/6-31G(2df,p) geometries and zero-point energies scaled by 0.9854, taking about 10%−15% more time than G3 due to a B3LYP/6-31G-(2df,p) frequency calculation. Besides the G3 theory, there are few other computational protocols for small molecules with high accuracy on the order of sub-kcal/mol: for instance, the "Wn theory" developed by Weizmann[18,19] and the highly accurate extrapolated ab initio thermochemistry (HEAT).[19] These two protocols show an unprecedented accuracy of 0.16 kcal/mol root-mean-square deviation between theory and experiment. This is less than the 0.6 kcal/mol deviation obtained with the G3x

---

[†] Department of Chemical Engineering.
[‡] Department of Electrical and Computer Engineering.

Precise Atomization Energies of CHON Molecules

J. Phys. Chem. A, Vol. 111, No. 43, 2007 **11161**

TABLE 1: Total Electronic Energies for Different Level of Theory

molecule	name	G3x (Ha)	B3PW91/6-31G** (Ha)	B3PW91/cc-pVTZ (Ha)	HF/6-31G* (Ha)	HF/3-21G (Ha)	MP2(full)/6-31G* (Ha)
				Training Set			
H (^{2}S)	hydrogen	−0.50097	−0.50218	−0.50398	−0.49823	−0.49620	−0.49823
C (^{3}P)	carbon	−37.82831	−37.82569	−37.83683	−37.68086	−37.48107	−37.73651
N (^{4}S)	nitrogen	−54.56490	−54.56393	−54.58043	−54.38544	−54.10539	−54.45945
O (^{3}P)	oxygen gas	−75.03224	−75.03133	−75.06125	−74.78393	−74.39366	−74.88200
H$_2$	hydrogen gas	−1.17724	−1.17752	−1.17894	−1.12683	−1.12296	−1.14414
CH (^{2}P)	methylidyne	−38.46536	−38.45947	−38.47372	−38.26493	−38.05191	−38.34240
CH$_2$ (^{3}B$_1$)	methylene	−39.13566	−39.13037	−39.15400	−38.92150	−38.70907	−39.00744
CH$_2$ (^{1}A$_1$)	methylene	−39.12003	−39.11058	−39.12739	−38.87237	−38.65185	−38.97401
CH$_3$ (^{2}A$_2$)	methyl radical	−39.82220	−39.82752	−39.84165	−39.55899	−39.34261	−39.67303
CH$_4$	methane	−40.50182	−40.50869	−40.52128	−40.19517	−39.97688	−40.33704
NH (^{3}S$_g$)	imidogen	−55.20079	−55.20039	−55.22008	−54.95942	−54.65924	−55.06142
NH$_2$ (^{2}B$_1$)	amino radical	−55.85778	−55.85708	−55.88010	−55.55770	−55.24538	−55.69375
NH3	ammonia	−56.54157	−56.53751	−56.56283	−56.18436	−55.87220	−56.35738
OH	hydroxyl radical	−75.70437	−75.69942	−75.73347	−75.38228	−74.97023	−75.52321
OH$_2$	water	−76.40430	−76.39255	−76.43068	−76.01075	−75.58596	−76.19924
HCCH	acetylene	−77.30429	−77.29452	−77.32507	−76.81783	−76.39596	−77.07622
H$_2$CCH$_2$	ethylene	−78.55861	−78.56139	−78.58780	−78.03172	−77.60099	−78.29429
H$_3$CCH$_3$	ethane	−79.79731	−79.80958	−79.83264	−79.22875	−78.79395	−79.50397
CN(2)	cyano radical	−92.68250	−92.66988	−92.70166	−92.20483	−91.68475	−92.44196
HCN	hydrogen cyanide	−93.39411	−93.38281	−93.41685	−92.87520	−92.35408	−93.16694
CO	carbon monoxide	−113.27498	−113.26028	−113.30509	−112.73788	−112.09330	−113.02818
HCO (^{2}A′)	formyl radical	−113.80663	−113.80451	−113.84911	−113.24766	−112.60380	−113.54033
H$_2$CO	formaldehyde	−114.45956	−114.45657	−114.50040	−113.86633	−113.22182	−114.17496
H$_3$COH	methyl alcohol	−115.68142	−115.68061	−115.72638	−115.03542	−114.39802	−115.35330
N$_2$	nitrogen gas	−109.49354	−109.47710	−109.51944	−108.94395	−108.30095	−109.26157
H$_2$NNH$_2$	hydrazine	−111.82992	−111.82834	−111.87306	−111.16937	−110.55001	−111.50440
NO(2)	nitric oxide	−129.84069	−129.83582	−129.88633	−129.24788	−128.50064	−129.56446
O$_2$ (^{3}S$_{gg}$)	oxygen gas	−150.25622	−150.26254	−150.32181	−149.61791	−148.76909	−149.95432
HOOH	hydrogen peroxide	−151.49339	−151.48602	−151.55190	−150.76479	−149.94582	−151.13492
CO$_2$	carbon dioxide	−188.51583	−188.50695	−188.58282	−187.63418	−186.56126	−188.11836
CH$_2$=C=CH$_2$	allene	−116.61032	−116.61529	−116.65394	−115.86110	−115.21991	−116.24852
CH$_3$CCH	propyne	−116.61218	−116.61088	−116.65105	−115.86432	−115.22539	−116.25624
HCOOCH$_3$	methyl formate	−228.97853	−228.97910	−229.06267	−227.78942	−226.51428	−228.43398
C$_3$H$_4$	cyclopropene	−116.57363	−116.58215	−116.61874	−115.82305	−115.16201	−116.21957
CH$_3$CH=CH$_2$	propylene	−117.86195	−117.87044	−117.90708	−117.07147	−116.42401	−117.46966
C$_3$H$_6$	cyclopropane	−117.84908	−117.86399	−117.89895	−117.05887	−116.40121	−117.46284
C$_3$H$_8$	propane	−119.09646	−119.11228	−119.14583	−118.26365	−117.61330	−118.67441
CH$_2$CHCHCH$_2$	butadiene	−155.93206	−155.93828	−155.98766	−154.91965	−154.05946	−155.44171
C$_4$H$_6$	2-butyne	−155.91857	−155.92530	−155.97503	−154.90925	−154.05365	−155.43515
C$_4$H$_6$	methylene cyclopropane	−155.90033	−155.91467	−155.96194	−154.88734	−154.01873	−155.41602
C$_4$H$_6$	bicyclobutane	−155.88818	−155.90576	−155.95085	−154.87177	−153.98664	−155.40948
C$_4$H$_6$	cyclobutene	−155.91224	−155.92595	−155.97140	−154.89962	−154.03072	−155.42933
C$_4$H$_8$	cyclobutane	−157.15113	−157.17061	−157.21373	−156.09720	−155.23136	−156.63706
C$_4$H$_8$	isobutene	−157.16684	−157.17888	−157.22567	−156.11067	−155.24714	−156.64640
C$_4$H$_{10}$	trans butane	−158.39583	−158.41496	−158.45895	−157.29841	−156.43247	−157.84497
C$_5$H$_8$	spiropentane	−195.19512	−195.21808	−195.27393	−193.91776	−192.82114	−194.58924
CH$_3$CN	methyl cyanide	−132.70716	−132.70395	−132.74809	−131.92753	−131.19180	−132.35131
CH$_3$ONO	methyl nitrite	−244.91551	−244.91763	−245.00711	−243.66864	−242.28636	−244.33911
HCOOH	formic acid	−189.69375	−189.68924	−189.76390	−188.76231	−187.70020	−189.25187
NCCN	cyanogen	−185.58943	−185.57208	−185.63549	−184.59122	−183.56292	−185.17464
(CH$_3$)$_2$NH	dimethylamine	−135.11378	−135.12415	−135.16693	−134.23885	−133.49485	−134.67810
CH$_2$CO	ketene	−152.54102	−152.54067	−152.59858	−151.72467	−150.87653	−152.16008
C$_2$H$_4$O	oxirane	−153.72969	−153.73575	−153.78991	−152.86736	−152.00070	−153.31569
CH$_3$CHO	acetaldehyde	−153.77197	−153.77548	−153.82981	−152.91597	−152.05525	−153.35897
HCOCOH	glyoxal	−227.73499	−227.72946	−227.81342	−226.59218	−225.30696	−227.20373
CH$_2$=CHCN	acrylonitrile	−170.76907	−170.76362	−170.82004	−169.76802	−168.82040	−170.31611
CH$_3$COOH	acetic acid	−229.00475	−229.00479	−229.08972	−227.81065	−226.53424	−228.43398
(CH$_3$)$_2$CHOH	isopropyl alcohol	−194.29116	−194.29720	−194.36323	−193.11542	−192.04826	−193.70655
(CH$_3$)$_3$N	trimethylamine	−174.40799	−174.42144	−174.47355	−173.26930	−172.31027	−173.84646
C$_4$H$_4$O	furan	−229.93288	−229.93998	−230.01669	−228.62521	−227.35008	−229.33278
C$_4$H$_5$N	pyrrole	−210.08961	−210.09891	−210.16667	−208.80785	−207.64757	−209.50418
C$_5$H$_5$N	pyridine	−248.18412	−248.19560	−248.27075	−246.69582	−245.31201	−247.51068
CCH (^{2}A′)	ethynyl radical	−76.58155	−76.57238	−76.60069	−76.15009	−75.73206	−76.35347
C$_2$H$_3$ (^{2}A′)	vinyl	−77.87039	−77.87411	−77.90104	−77.39029	−76.96280	−77.61326
CH$_3$CO (^{2}A′)	acetyl radical	−153.11880	−153.12322	−153.17847	−152.29798	−151.43851	−152.72265
H$_2$COH (^{2}A)	hydroxymethyl radical	−115.01578	−115.01734	−115.06394	−114.40876	−113.77382	−114.70340
C$_2$H$_5$ (^{2}A′)	ethyl radical	−79.12299	−79.13604	−79.16052	−78.59715	−78.16365	−78.84466
(CH$_3$)$_2$CH (^{2}A′)	isopropyl radical	−118.42610	−118.44521	−118.47991	−117.63614	−116.98609	−118.01923
(CH$_3$)$_3$C (^{2}A$_1$)	*tert*-butyl radical	−157.73080	−157.75412	−157.79892	−156.67501	−155.80910	−157.19579
NO$_2$ (2)	nitrogen dioxide	−204.99262	−204.99491	−205.07605	−204.03149	−202.84346	−204.56859
C$_4$H$_6$	methylallene	−155.91220	−155.92227	−155.97108	−154.89938	−154.04148	−155.42290
C$_5$H$_8$	isoprene	−195.23656	−195.24541	−195.30490	−193.95723	−192.87756	−194.61762

TABLE 1: Continued

molecule	name	G3x (Ha)	B3PW91/ 6-31G** (Ha)	B3PW91/ cc-pVTZ (Ha)	HF/6-31G* (Ha)	HF/3-21G (Ha)	MP2(full)/ 6-31G* (Ha)
C_5H_{10}	cyclopentane	−196.48266	−196.50359	−196.55724	−195.16358	−194.08847	−195.84010
C_5H_{12}	*n*-pentane	−197.69520	−197.71759	−197.77198	−196.33310	−195.25156	−197.01553
C_5H_{12}	neopentane	−197.70229	−197.71910	−197.77325	−196.33382	−195.25674	−197.02326
C_6H_8	1,3-cyclohexadiene	−233.33068	−233.34346	−233.41187	−231.83190	−230.54323	−232.62282
C_6H_8	1,4-cyclohexadiene	−233.33039	−233.34313	−233.41223	−231.83358	−230.54358	−232.62087
C_6H_{12}	cyclohexane	−235.79261	−235.81539	−235.87841	−234.20801	−232.91691	−235.02161
C_6H_{14}	*n*-hexane	−236.99460	−237.02021	−237.08502	−235.36779	−234.07067	−236.18612
C_6H_{14}	3-methylpentane	−236.99550	−237.01801	−237.08262	−235.36430	−234.06936	−236.18660
$C_6H_5CH_3$	toluene	−271.45783	−271.47487	−271.55409	−269.74017	−268.24021	−270.66287
C_7H_{16}	*n*-heptane	−276.29402	−276.32283	−276.39806	−274.40249	−272.88977	−275.35670
C_8H_8	cyclooctatetraene	−309.46732	−309.47438	−309.56776	−307.52422	−305.80706	−308.56609
C_8H_{18}	*n*-octane	−315.59346	−315.62546	−315.71111	−313.43718	−311.70888	−314.52729
$C_{10}H_8$	naphthalene	−385.73775	−385.75680	−385.86847	−383.35505	−381.21581	−384.66461
$C_{10}H_8$	azulene	−385.68030	−385.70195	−385.81399	−383.28262	−381.13780	−384.45974
CH_3COOCH_3	acetic acid methyl ester	−268.28912	−268.29377	−268.38764	−266.83683	−265.34743	−267.59323
$(CH_3)_3COH$	*tert*-butyl alcohol	−233.59722	−233.60388	−233.67969	−232.15347	−230.87376	−232.88475
$C_6H_5NH_2$	aniline	−287.49496	−287.50748	−287.59724	−285.73082	−284.14352	−286.68343
C_6H_5OH	phenol	−307.35194	−307.36112	−307.46161	−305.55806	−303.86010	−306.52345
C_4H_6O	divinyl ether	−231.10928	−231.11375	−231.19359	−229.76072	−228.48803	−230.46190
C_4H_8O	tetrahydrofuran	−232.36337	−232.37635	−232.44996	−230.97645	−229.69916	−231.69229
C_5H_8O	cyclopentanone	−270.47191	−270.48618	−270.57010	−268.86641	−267.36810	−269.71011
$C_6H_4O_2$	benzoquinone	−381.30584	−381.30682	−381.43360	−379.23557	−377.10068	−380.36128
$C_4H_4N_2$	pyrimidine	−264.21714	−264.22580	−264.30741	−262.69349	−261.20619	−263.53629
$N≡C−CH_2−CH_2−C≡N$	butanedinitrile	−264.20900	−264.19766	−264.28339	−262.69024	−261.22345	−263.53383
$C_4H_4N_2$	1,4-dipyridine	−264.20975	−264.21907	−264.30056	−262.68301	−261.19750	−263.53042
$CH_3−C(=O)−CCH$		−229.88931	−229.88518	−229.96494	−228.58911	−227.31173	−229.28772
$CH_3−CH=CH−CHO$	crotonaldehyde	−231.14461	−231.15181	−231.22875	−229.80397	−228.51499	−230.50614
$CH_3−C(=O)−O−$ $C(=O)−CH_3$	acetic anhydride	−381.58694	−381.58938	−381.72352	−379.58350	−377.45878	−380.64634
$(CH_3)_2CH−CN$	2-methylpropanenitrile	−211.30738	−211.30972	−211.37448	−209.99800	−208.83265	−210.69465
$CH_3−CO−CH_2−CH_3$	methyl ethyl ketone	−232.38239	−232.39412	−232.46862	−230.99805	−229.70935	−231.71266
$(CH_3)_2CH−CHO$	2-methylpropanal	−232.37222	−232.38116	−232.45612	−230.98580	−229.69708	−231.70196
$C_4H_8O_2$	1,4-dioxane	−307.54636	−307.55383	−307.65637	−305.82531	−304.12996	−306.71829
C_4H_8NH	tetrahydropyrrole	−212.50635	−212.52152	−212.58470	−211.14533	−209.97508	−211.85005
$CH_3−CH_2−CH(CH_3)−NO_2$	nitro-*sec*-butane	−362.82890	−362.83904	−362.96146	−360.77464	−358.72770	−361.86967
$CH_3−CH_2−O−CH_2−CH_3$	diethyl ether	−233.57537	−233.58947	−233.66413	−232.14489	−230.86173	−232.86743
$CH_3−CH(OCH_3)_2$	dimethylacetal	−308.75986	−308.76727	−308.87086	−306.98426	−305.28599	−307.88511
$(CH_3)_3C−NH_2$	*tert*-butylamine	−213.73552	−213.74557	−213.81004	−212.32183	−211.15083	−213.04083
cyc-CH=CH−N(CH_3)− CH=CH−	*n*-methylpyrrole	−249.38316	−249.39621	−249.47347	−247.83878	−246.46132	−248.67238
$C_5H_{10}O$	tetrahydropyran	−271.67077	−271.68595	−271.76875	−270.01791	−268.52560	−270.87133
$CH_3−CH_2−CO−CH_2−CH_3$	diethyl ketone	−271.68242	−271.69769	−271.78222	−270.03365	−268.53118	−270.88427
$C_5H_{10}O_2$	isopropyl acetate	−346.89843	−346.90884	−347.02272	−344.91508	−342.99695	−345.94628
cyc-$C_5H_{10}NH$	piperidine	−251.81419	−251.83231	−251.90408	−250.18871	−248.80326	−251.03088
$(CH_3)_2CH−O−CH(CH_3)_2$	diisopropyl ether	−312.18341	−312.20063	−312.29483	−310.21885	−308.50855	−311.21918
C_6H_5 $(^2A_1)$	phenyl radical	−231.46083	−231.47942	−231.54862	−230.06507	−228.78331	−230.77775
$(CH_3)_3C−O−CH_3$	*tert*-butyl methyl ether	−272.87831	−272.89010	−272.97415	−271.17563	−269.68338	−272.04192
C_6H_6	benzene	−232.15434	−232.16805	−232.23735	−230.70314	−229.41945	−231.48719
CH_3NO_2	nitromethane	−244.91970	−244.92208	−245.01440	−243.66199	−242.25586	−244.34534
Validation Set							
C_4H_{10}	isobutane	−158.39842	−158.41577	−158.45971	−157.29898	−156.43447	−157.84777
CH_3NH_2	methylamine	−95.82440	−95.82869	−95.86254	−95.20983	−94.68166	−95.51444
CH_3CONH_2	acetamide	−209.14096	−209.14471	−209.21952	−207.97601	−206.81580	−208.58499
$CH_3CH_2NH_2$	ethylamine	−135.12630	−135.13510	−135.17909	−134.24761	−133.50415	−134.68825
CH_3COCH_3	acetone	−193.08233	−193.09041	−193.15482	−191.96224	−190.88722	−192.54087
C_2H_4NH	aziridine	−133.87226	−133.88157	−133.92515	−133.03856	−132.27801	−133.47309
$C_2H_5OCH_3$	methoxyethane	−194.27102	−194.28148	−194.34578	−193.10487	−192.03755	−193.69148
CH_3CH_2OH	ethanol	−154.98554	−154.98883	−155.04506	−154.07574	−153.22268	−154.52896
CH_3OCH_3	dimethyl ether	−154.96662	−154.97336	−155.02727	−154.06475	−153.21321	−154.51555
CH_3CH_2O $(^2A'')$	ethoxy radical	−154.30395	−154.31847	−154.37124	−153.46094	−152.61640	−153.86706

theory. Certainly the set of molecules with such small tolerance is very limited.

Comparing theoretical and experimental energies requires of a mixture of corrections to perform a fair comparison, as both are relative energies with respect to different references at different conditions. Theoretical energies are obtained at 0 K in vacuum having as reference the individual atoms composing the molecule; however, the experimental energies are taken at ambient conditions of pressure and temperature with respect to reference states that can be simple gas molecules, liquids, or solids depending on each particular case.

The atomization energy of a molecule is defined as

$$D_o = \left(\sum_{i=atoms} E_i \right) - E_{molecule} - E_{ZPE} \qquad (1)$$

where E_i and $E_{molecule}$ are the individual energies of all atoms in the molecule and the total energy of the molecule, respec-

Precise Atomization Energies of CHON Molecules

J. Phys. Chem. A, Vol. 111, No. 43, 2007 **11163**

tively, and E_{ZPE} is the zero-point vibrational energy correction of the molecule. Thus D_o is the energy stabilization due to the formation of the molecule from the participant atoms. Table A1 in the Supporting Information reports the atomization energy of the G3/99 set for different levels of theory using the above formula.

On the other hand, the enthalpy of formation of a molecule is defined as

$$\Delta_f H_0{}^\circ{}_{molecule} = \Delta_f H_{298}{}^\circ - (H_{298}{}^\circ - H_0{}^\circ)_{molecule} + \sum_{atoms} (H_{298}{}^\circ - H_0{}^\circ) \quad (2)$$

where $(H_{298}{}^\circ - H_0{}^\circ)_{atoms}$ experimental values for H, C, O, and N are 1.01, 0.25, 1.04, and 1.04 kcal/mol, respectively.

$$(H_{298}{}^\circ - H_0{}^\circ)_{molecule} = H^\circ corr - \epsilon_{ZPE} \quad (3)$$

where $H^\circ corr - \epsilon_{ZPE}(M)$ is calculated for each method.

And then, the experimental atomization energy, as suggested by Curtiss,[17] is calculated using the following formulas:

$$D_o{}^{Experimental} = (\sum_{atoms} \Delta_f H_0{}^\circ) - \Delta_f H_0{}^\circ{}_{molecule} \quad (4)$$

where $\Delta_f H_0{}^\circ{}_{atoms}$ experimental for H, C, O, and N are 51.63, 169.98, 58.99, and 112.53 kcal/mol, respectively.

III. Effect of the Correctors

The calculation of the experimental atomization energy following this procedure is reported in Table A1 (Supporting Information) for all molecules; then the average error for a level of theory is obtained from the mean absolute error average of the atomization energy with respect to experimental values. All calculations are performed using the program Gaussian-03,[20] Table 1 reports the total electronic energies of the G3/99 set for different levels of theory. Table A2 (Supporting Information) reports the zero point energy corrections calculated using their corresponding level of theory for the calculation of their optimized geometries and energies. The results for the training and validation set, with and without the correctors described in the next section, are reported in Tables 3 and 4, respectively.

Several correctors are used in this work, and as in our previous work, we introduce a multiplicative factor that compensate for the deficiencies to predict the total electronic energy. Then, the corrected total energy ($E^{corrected}$) for each atom or molecule is given by

$$E^{corrected} = \epsilon_E E + E_{ZPE}$$
corrected total energy using the multiplicative factor (5)

This multiplicative factor allows us to make proportional changes to the energy of each molecule. However, to improve the atomization energies, we introduce corrective factors such that they minimize the average errors for a specific level of theory (method/basis-set). Complex bond additive correction procedures were developed since the 1980s to reduce the demand of computational resources or improve accuracy,[21] instead we use even simpler additive factors ΔB_i, called boundary factors, that compensate for the deficiencies to predict the bond energies. We consider a different factor for each type

TABLE 2: Corrector Factors for the Training Set (115 Molecules)[a]

(a) only B_i	B_0	B_1	B_2	B_3
G3x	−0.5	−0.5	−0.6	−0.8
B3PW91/6-31G**	2.4	3.1	3.7	0.4
B3PW91/cc-pvtz	1.8	1.7	4.3	2.2
HF/6-31G*	20.0	−4.1	−32.1	−51.8
HF/3-21G	29.4	2.8	−37.8	−53.6
MP2(full)/6-31G*	11.6	4.7	9.9	13.2

(b) B_i and ϵ	B_0	B_1	B_2	B_3	ϵ
G3x	−0.5	−0.7	−0.9	−1.2	0.998
B3PW91/6-31G**	2.3	2.6	3.0	−0.6	0.995
B3PW91/cc-pvtz	1.8	2.6	5.7	4.0	1.009
HF/6-31G*	17.7	−34.9	−75.9	−104.8	0.613
HF/3-21G	22.9	−43.6	−93.3	−126.0	0.440
MP2(full)/6-31G*	12.9	−2.4	−2.5	−2.9	0.909

(c) ΔE_i	ΔE_4	ΔE_3	ΔE_2	ΔE_1
G3x	−0.1	0.0	0.0	0.1
B3PW91/6-31G**	0.5	−1.3	−0.7	−4.5
B3PW91/cc-pvtz	−1.2	−1.4	0.7	−1.9
HF/6-31G*	61.8	47.7	11.7	−2.4
HF/3-21G	78.2	57.1	10.4	−10.6
MP2(full)/6-31G*	8.0	2.0	9.4	−0.7

(d) ΔE_i, ϵ	ΔE_4	ΔE_3	ΔE_2	ΔE_1	ϵ
G3x	−1.0	−1.3	−0.6	−0.3	1.009
B3PW91/6-31G**	5.8	5.3	2.8	−1.6	0.955
B3PW91/cc-pvtz	3.8	5.5	4.1	1.3	0.955
HF/6-31G*	70.3	59.8	19.5	5.2	0.880
HF/3-21G	88.5	80.1	24.9	6.4	0.767
MP2(full)/6-31G*	13.4	8.7	12.7	1.4	0.952

(e) ϵ	ϵ
G3x	1.000
B3PW91/6-31G**	0.994
B3PW91/cc-pvtz	1.000
HF/6-31G*	1.357
HF/3-21G	1.421
MP2(full)/6-31G*	1.073

(f) atom factors	A_h	A_c	A_n	A_o
G3x	−0.1	0.2	0.1	−0.5
B3PW91/6-31G**	0.0	−2.0	1.1	1.2
B3PW91/cc-pvtz	1.2	−2.2	−1.1	0.6
HF/6-31G*	12.6	46.1	62.0	51.8
HF/3-21G	12.0	54.7	82.3	72.0
MP2(full)/6-31G*	9.7	1.5	7.7	2.6

(g) atom factors and ϵ	A_h	A_c	A_n	A_o	ϵ
G3x	−0.1	0.2	0.1	−0.5	1.000
B3PW91/6-31G**	1.4	1.8	3.3	3.4	0.997
B3PW91/cc-pvtz	0.5	−3.9	−2.1	−0.6	1.011
HF/6-31G*	11.3	43.2	60.0	50.0	1.020
HF/3-21G	12.2	55.3	82.7	72.4	0.996
MP2(full)/6-31G*	15.0	13.8	15.5	10.1	0.927

[a] All quantities are in kcal/mol except for ϵ, which is unitless.

of bond; thus, the corrected total energy ($E^{corrected}$) for each atom or molecule is given by

$$E^{corrected} = E + E_{ZPE} + \Delta E_{bound}$$
corrected total energy using the boundary factors (6)

where $\Delta E_{bound} = \sum_{i=0}^{3} B_i * N_i$; B_0, B_1, B_2, and B_3 are factors for no bond, single, double, and triple bonds; and N_i is the number of bonds of type i.

This means that for the atoms C, H, O, and N, B_0 equals 1, and for the molecules B_0 equals zero. Because our test of molecules is dominated by single bonds, our factors B_0 and B_1 yield changes proportional to the number of atoms or the energy

TABLE 3: Average Errors Using the Training Set for Several Levels of Theory

level of theory	no correctors	B_i	$B_i \& \epsilon$	ΔE_i	$\Delta E_i \& \epsilon$	ϵ	atomic factors	atom factors $\& \epsilon$
G3x	0.7	0.7	0.7	0.7	0.7	0.7	0.6	0.6
B3PW91/6-31G**	6.4	2.1	2.1	2.6	2.1	3.0	2.4	2.4
B3PW91/cc-pVTZ	3.1	1.9	1.9	2.6	2.0	3.1	2.1	2.0
HF/6-31G*	271.2	20.1	14.8	6.8	4.7	37.3	6.8	6.7
HF/3-21G	313.5	30.2	17.3	12.4	5.7	57.9	7.1	7.1
MP2(full)/6-31G*	66.3	5.6	4.9	5.6	5.2	13.0	4.4	3.8

TABLE 4: Average Errors Using the Validation Set for Several Levels of Theory

level of theory	no correctors	B_i	$B_i \& \epsilon$	ΔE_i	$\Delta E_i \& \epsilon$	ϵ	atom factors	atom factors $\& \epsilon$
G3x	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
B3PW91/6-31G**	3.9	1.8	1.8	1.2	1.4	2.2	1.3	1.5
B3PW91/cc-pvtz	1.9	1.4	1.3	1.0	1.3	1.8	1.3	1.3
HF/6-31G*	241.0	13.7	8.1	4.5	3.1	23.4	5.2	5.2
HF/3-21G	276.1	18.2	9.6	6.6	4.7	37.2	6.1	6.1
MP2(full)/6-31G*	72.5	4.7	3.3	4.9	3.7	14.6	3.7	2.8

of each molecule. Also, as in our previous work, we introduce additive factors $\Delta E_{\text{multiplicity}}$, such as ΔE_1, ΔE_2, ΔE_3, and ΔE_4, for singlet, doublet, triplet, and quartet systems, respectively, to compensate for the errors in the multiplicity or total spin of the atoms and molecules. Considering these correctors, the corrected total energy ($E^{\text{corrected}}$) for each atom or molecule is given by

$$E^{\text{corrected}} = E + E_{\text{ZPE}} + \Delta E_{\text{multiplicity}}$$

corrected total energy using the multiplicity factors (7)

Due to the characteristics of our test set (most of the molecules are spin singlets), the factor ΔE_1 represents an additive factor to all the molecules, independently of the magnitude of its energy. Also, notice ΔE_4 represents an additive factor to the energy of nitrogen, ΔE_2 is an additive factor to the energy of hydrogen, and ΔE_3 is an average corrector among carbon and oxygen. These last three multiplicity factors yield also a multiplicative nature because their corrections increase as the number of atoms increases.

We now introduce additive factors to the energy of the four elements of our set of molecules, A_c, A_h, A_o, A_n, further simplifying the corrections. Considering this, just the energy of each atom changes to

$$E_i^{\text{corrected}} = E_i + A_i$$

corrected total atomic energy using the atomic factors (8)

and the energy of the molecules remains unchanged. These atomic factors are expected to be equally as effective as the multiplicity factors, because these also share an additive and multiplicative nature and are even more specific to the correction of the energies of carbon and oxygen.

Analyzing only the multiplicative factor, ϵ, from Table 2e, we can have an idea of the quality of each level of theory (method/basis). Values close to 1 would correspond to the more exact methods, and distant values from 1 would correspond to poor methods. We can rank our additive factors by comparing the multiplicative factor ϵ. Thus, the most effective factors are the atomic factors, followed by the multiplicity factors, and the boundary factors. The multiplicative factor also allows us to test the stability of our corrective factors; the smaller the change of each type of additive correctors when combined with multiplicative corrector (ϵ), the better the stability. These changes are shown in Table 2 sections b, d, and g. This stability

is also proved in the validation set (Table 4), showing that the atomic factors are the most robust.

There is a similarity between the multiplicity and atomic factors, due to the characteristics of our set of molecules. Most of the molecules in the training set have multiplicity 1; therefore, the factors ΔE_2, ΔE_3, and ΔE_4 yield corrections to the energies of hydrogen, carbon, and hydrogen, respectively. The relationship between these factors is the same in both types of corrections. Moreover, although the difference among the atomic factors of carbon and oxygen is larger, the accuracy using atomic factors is better than using multiplicity factors.

The source of the error for the MP2 (full)/6-31G* level of theory is mainly due to the prediction of the hydrogen and oxygen energies (Table 2 section f). Comparing the 6-31G** and cc-pVTZ basis sets, we notice that the hydrogen energy needs stronger corrections when an extended basis is used. The 6-31G** and cc-pvtz basis sets underestimate and overestimate, respectively, the nitrogen energy, and the error in the calculation for the carbon energy is the same with both. We can conclude that the accuracy of cc-pVTZ is due mostly to the calculation for the oxygen energy. Moreover, comparing the HF/6-31G* and HF/3-21G levels of theory, we also notice that the hydrogen energy needs stronger corrections when an extended basis set is used. The atomic corrections increase quasi-linearly with the energy, in agreement with the results in Table 2 section f. We can conclude observing the results in Table 3 that the best correction is the one when the atom energies are corrected by an additive energy, not only for simplicity and accuracy but also for stability. These implies only four parameters need to be include on the atomic energies; the molecular energies are untouched by these corrections.

Tables 3 and 4 show that the average error without corrections is smaller in the validation set than in the training set, because our training set contains more large molecules than small ones. Then, because our methodology is size consistent, it is expected that the corrected absolute errors in all the methods be larger for a set of larger molecules, as shown in Tables 3 and 4. However, the relative corrections (per bond, atom, or percentage) are almost the same in both sets, which means, for instance, that the error per bond is maintained as the size increases.

An important detail of the atomic factors is that their corrections are better for bigger molecules. For instance, the average error for the first 25 low-energy molecules (the small molecules of the training set) yields average errors of 6.5 kcal/mol, and the average error for the 25 first high-energy molecules (the large molecules of the training set) is 2.5 kcal/mol. This is

Precise Atomization Energies of CHON Molecules

J. Phys. Chem. A, Vol. 111, No. 43, 2007 **11165**

because our methodology takes advantage of the fact that, as the energies of the atoms are already extremely well-corrected, the larger systems benefit because the larger basis set helps each individual atom to reshape their structure much better than when a smaller number of atoms (thus smaller basis set) is used. To validate the use of the reported energy correctors, we use them in 10 representative molecules from the G3 set shown in Table 1 that are not used in the training set; the results are shown in Table 4, which is practically self-explanatory, as it shows similar and better results than those obtained with the training set. The selection of just 10 molecules (compared with the 115 molecules for the training set) proves the specificity of these correctors. Certainly the G3x cannot be improved further because all corrections imposed on the methods are in some way considered in the correctors. Having the B3PW91 with errors of only 1.5 and 1.3 with the small 6-31G* and the cc-pVTZ basis sets, respectively, is really encouraging for the application of these methods to large systems.

IV. Conclusions

We report correctors for the total energies of a few levels of theory commonly used in quantum chemistry calculations. It is found that these correctors can make strong effects on the accuracy of practically all methods and that the strategic evaluation of additive and multiplicative correctors provides a good description of the quality of the correctors. The most practical and easiest of the corrections is simply to correct the atom energies by an additive corrector. For instance, for the popular HF/6-31G* method the error in the atomization energies is on average 271.2 kcal/mol, which is reduced to only 6.7 kcal/mol when the energy of the CHON atoms are corrected by additive energies. This correction certainly makes the HF/6-31G* level of theory very useful. On the other hand, the B3PW91/cc-pVTZ yield corrected average errors of 2.0 kcal/mol, which are already within chemical accuracy not even attainable by any of the uncorrected standard ab initio methods. These results justify the use of ab initio energy differences where the energy of the atoms is cancelled and we hope this work triggers its extension to molecules with other atoms in the periodic table.

Note Added after ASAP Publication. This Article was posted ASAP on October 10, 2007. The headings for the second columns in Tables 3 and 4 were changed from "no. of correctors" to "no correctors". The corrected Article was reposted on October 15, 2007.

Supporting Information Available: Tables of atomization energies and zero point energy corrections. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Seminario, J. M.; Maffei, M. G.; Agapito, L. A.; Salazar, P. F. *J. Phys. Chem. A* **2006**, *110*, 1060.

(2) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90*, 5622.

(3) Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1990**, *93*, 2537.

(4) Duan, X. M.; Song, G. L.; Li, Z. H.; Wang, X. J.; Chen, G. H.; Fan, K. N. *J. Chem. Phys.* **2004**, *121*, 7086.

(5) Cioslowski, J.; Schimeczek, M.; Liu, G.; Stoyanov, V. *J. Chem. Phys.* **2000**, *113*.

(6) Agapito, L. A.; Maffei, M. G.; Salazar, P. F.; Seminario, J. M. *J. Phys. Chem. A* **2006**, *110*, 4260.

(7) Seminario, J. M. *Int. J. Quantum Chem.* **1994**, *S28*, 655.

(8) Seminario, J. M. *Chem. Phys. Lett.* **1993**, *206*, 547.

(9) Wang, X.; Wong, L.; Hu, L.; Chan, C.; Su, Z.; Chen, G. *J. Phys. Chem A* **2004**, *108*, 8514.

(10) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764.

(11) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374.

(12) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2005**, *123*, 124107.

(13) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(14) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(15) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244.

(16) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671.

(17) Larry, A. C.; Paul, C. R.; Krishnan, R.; John, A. P. *J. Chem. Phys.* **2001**, *114*, 108.

(18) Karton, A.; Ruscic, B.; Martin, J. M. L. *J. Mol. Struct.*: THEOCHEM **2007**, *811*, 345.

(19) Tajti, A.; Szalay, P. G.; Csaszar, A. G.; Kallay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vazquez, J.; Stanton, J. F. *The Journal of Chemical Physics* **2004**, *121*, 11599.

(20) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(21) Zachariah, M. R.; Melius, C. F. *J. Phys. Chem. A* **1997**, *101*, 913.