

# An improved theoretical approach to the empirical corrections of density functional theory

Jenn-Huei Lii · Ching-Han Hu

Received: 31 August 2011 / Accepted: 13 December 2011 / Published online: 24 December 2011  
© Springer Science+Business Media B.V. 2011

**Abstract** An empirical correction to density functional theory (DFT) has been developed in this study. The approach, called correlation corrected atomization–dispersion (CCAZD), involves short- and long-range terms. Short-range correction consists of bond ( $I,2$ -) and angle ( $I,3$ -) interactions, which remedies the deficiency of DFT in describing the proto-branching stabilization effects. Long-range correction includes a Buckingham potential function aiming to account for the dispersion interactions. The empirical corrections of DFT were parameterized to reproduce reported  $\Delta H_f$  values of the training set containing alkane, alcohol and ether molecules. The  $\Delta H_f$  of the training set molecules predicted by the CCAZD method combined with two different DFT methods, B3LYP and MPWB1K, with a 6-31G\* basis set agreed well with the experimental data. For 106 alkane, alcohol and ether compounds, the average absolute deviations (AADs) in  $\Delta H_f$  were 0.45 and 0.51 kcal/mol for B3LYP- and MPWB1K-CCAZD, respectively. Calculations of isomerization energies, rotational barriers and conformational energies further validated the CCAZD approach. The isomerization energies improved significantly with the CCAZD treatment. The AADs for 22 energies of isomerization reactions were decreased from 3.55 and 2.44 to 0.55 and 0.82 kcal/mol for B3LYP and MPWB1K, respectively. This study also provided predictions of MM4, G3, CBS-QB3 and B2PLYP-D for comparison. The final test of the

CCAZD approach on the calculation of the cellobiose analog potential surface also showed promising results. This study demonstrated that DFT calculations with CCAZD empirical corrections achieved very good agreement with reported values for various chemical reactions with a small basis set as 6-31G\*.

**Keywords** Density functional theory · Enthalpy of formation · CCAZ · CCAZD · MM4 · G3 · CBS-QB3 · B2PLYP-D

## Introduction

Density functional theories (DFTs) have become increasingly popular in electronic structure calculations for large systems due to their cost-effectiveness. Among those popular DFT methods, the hybrid functional B3LYP is regarded as a standard method for the study of organic molecules in the gas phase [1]. A nice thing about DFT is that simple approximations do remarkably well for the prediction of structures and thermodynamic properties of molecules. However, the form of exchange–correlation functional in the DFT formalism remains unknown and needs to be approximated. Although researchers have introduced better functionals since DFT initially gained popularity, the known problems of DFT in describing proto-branching stabilization effects [2] and long-range dispersion energies [3–6] more or less remains. Correcting these problems within the standard DFT framework seems to be very difficult, since all non-empirical attempts ultimately result with methods as complex as the simplest wave-function methods [7]. Over the past decade, many reports [1, 7–20] have discussed the application of certain practical approaches to fix the known DFT problems by

**Electronic supplementary material** The online version of this article (doi:10.1007/s10822-011-9534-x) contains supplementary material, which is available to authorized users.

J.-H. Lii · C.-H. Hu (✉)  
Department of Chemistry, National Changhua University  
of Education, Changhua 50058, Taiwan  
e-mail: chingkh@cc.ncue.edu.tw

adding an  $R^{-6}$  term to the usual DFT energy, where  $R$  represent the inter-atomic distances. Although these approaches have yielded some encouraging results, there is still plenty of room for improvement [21]. These approaches apply empirical corrections primarily to typical weakly bound molecules (inter-molecular interactions), and assume they are also applicable to intra-molecular interactions. However, Schleyer et al. [17], demonstrated that inter- and intra-molecular interactions are significantly different, and must be handled separately to obtain improved results. According to Schleyer's approach, the proto-branching stabilization and long-range dispersion problems of DFT are corrected by the intra- and inter-molecular corrections, respectively.

This study presents a theoretical approach, called correlation corrected atomization-dispersion (CCAZD), to the empirical corrections of DFT. Instead of dividing the empirical corrections into intra- and inter-molecular categories, this approach uses an MM4-type [22] dispersion potential to correct the long-range dispersion problem, and uses  $I,2$ - and  $I,3$ -corrections to correct the proto-branching stabilization problem. The DFT energy with the CCAZD correction  $E_{\text{DFT-CCAZD}}$  is thus the sum of usual DFT energy  $E_{\text{DFT}}$ , and all empirical corrections are as follows:

$$E_{\text{DFT-CCAZD}} = E_{\text{DFT}} + \sum E^{1,2} + \sum E^{1,3} + \sum E_{\text{DISP}}^{\geq 1,4}$$

where, the  $I,2$ -correction  $E^{1,2}$  is the interaction between two atoms separated by a bond; the  $I,3$ -correction  $E^{1,3}$  is the interaction between two atoms separated by two bonds; and the  $I,4$ - and higher correction  $E_{\text{DISP}}^{\geq 1,4}$  is the interaction between two inter-molecular atoms or two atoms separated by three bonds or more. In CCAZD approach, both  $E^{1,2}$  and  $E^{1,3}$  corrections are treated as group corrections, while  $E_{\text{DISP}}^{\geq 1,4}$  correction is a MM4-type dispersion potential. All empirical corrections were parameterized based on experimental enthalpies of formation. This study adopts a molecular mechanics parameterization strategy by starting with alkanes, alcohols and ethers before extending the procedure to other functional groups later. This study demonstrates that DFT calculations with CCAZD empirical corrections agree well with reported values for various chemical reactions, even with as small a basis set as 6-31G\*.

## Computational approach

### Enthalpies of formation ( $\Delta H_f$ )

The  $\Delta H_f$  of 56 alkanes, and of 54 alcohols and ethers were computed using atomization (AZ) and isodesmic [23] (ISOD) reactions. The full list of alkanes, alcohols, and

ethers used in the  $\Delta H_f$  calculations appears in the Tables 1S and 2S (Supporting Information). Allinger's MM4 [22], Pople's G3 [24], Petersson's CBS-QB3 [25], and Grimme's B2PLYP-D [14, 26] theories were applied, along with two DFT methods (B3LYP [27, 28] and MPWB1K [29]). The B2PLYP-D theory with a TZVP basis set, and B3LYP (B3) and MPWB1K (MP) theories with a 6-31G\* basis set were applied throughout all calculations to obtain optimized geometries and harmonic vibration frequencies. The vibration frequencies were scaled by 0.96, and were subsequently used to obtain zero-point-energy (ZPE) and thermal corrections to 298 K. The scale factor of 0.96 was chosen as it was used in the G3MP2B3 approach [30]. For B3LYP and MPWB1K theories, the  $\Delta H_f$  were also calculated with atomic equivalent (AE), group equivalent (GE) and the proposed CCAZD methods. All computations, except those with B2PLYP-D theory, were performed using the Gaussian03 program [31]. The Gaussian09 program was used for B2PLYP-D calculations.

### AZ and AE methods

The atomization method (AZ) has been widely used in semi-empirical quantum mechanics. This approach computes the enthalpy of formation of the molecule  $\Delta H_f$  using the following general equation:

$$\Delta H_f = H^{\text{mol}} - \sum_A H^A + \sum_A \Delta H_f^A \quad (1)$$

where  $H^{\text{mol}}$  is enthalpy of the molecule;  $H^A$  and  $\Delta H_f^A$  are enthalpy and enthalpy of formation of the atoms, respectively. Equation 1 can be expanded to Eqs. 2 and 3, for alkanes and alcohols/ethers, respectively.

$$\Delta H_f(\text{C}_n\text{H}_m) = H(\text{C}_n\text{H}_m) - n[H(\text{C}) - \Delta H_f(\text{C})] - m[H(\text{H}) - \Delta H_f(\text{H})] \quad (2)$$

$$\Delta H_f(\text{C}_n\text{H}_m\text{O}_x) = H(\text{C}_n\text{H}_m\text{O}_x) - n[H(\text{C}) - \Delta H_f(\text{C})] - m[H(\text{H}) - \Delta H_f(\text{H})] - x[H(\text{O}) - \Delta H_f(\text{O})] \quad (3)$$

The atomic equivalent (AE) method, conversely, originates from the AZ scheme. The AE approach replaces the terms  $[H(\text{C}) - \Delta H_f(\text{C})]$ ,  $[H(\text{H}) - \Delta H_f(\text{H})]$  and  $[H(\text{O}) - \Delta H_f(\text{O})]$  in atomization reactions (Eqs. 2, 3) with atomic equivalents,  $\varepsilon_{\text{C}}$ ,  $\varepsilon_{\text{H}}$  and  $\varepsilon_{\text{O}}$ , respectively (Eqs. 4, 5). These were thus least-squares fitted to reproduce the experimental  $\Delta H_f$  values of the training-set molecules. Table 1 summarizes the optimized atomic equivalents,  $\varepsilon_{\text{C}}$ ,  $\varepsilon_{\text{H}}$  and  $\varepsilon_{\text{O}}$ .

$$\Delta H_f(\text{C}_n\text{H}_m) = H(\text{C}_n\text{H}_m) - n\varepsilon_{\text{C}} - m\varepsilon_{\text{H}} \quad (4)$$

$$\Delta H_f(\text{C}_n\text{H}_m\text{O}_x) = H(\text{C}_n\text{H}_m\text{O}_x) - n\varepsilon_{\text{C}} - m\varepsilon_{\text{H}} - x\varepsilon_{\text{O}} \quad (5)$$

**Table 1** The optimized equivalents used in the AE and GE schemes (in hartrees)

	B3LYP/6-31G*		MPWB1K/6-31G*	
	Alkanes	Alcohols/ethers	Alkanes	Alcohols/ethers
AE				
$\varepsilon_C$	−38.1110274	−38.117157	−38.102423	−38.107579
$\varepsilon_H$	−0.583073	−0.579711	−0.574848	−0.572217
$\varepsilon_O$		−75.152922		−75.124401
GE				
$\varepsilon_{CH_4}$	−40.449169	−40.443691	−40.405823	−40.402341
$\varepsilon_{C_2H_4}$	−78.560431	−78.555302	−78.508846	−78.505484
$\varepsilon_{C_2H_6}$	−79.724267	−79.719982	−79.657141	−79.654151
$\varepsilon_{CH_3OH}$		−115.585580		−115.518192
$\varepsilon_{CH_3OCH_3}$		−154.872714		−154.778423
$\varepsilon_{H_2O}$		−76.292740		−76.253883

### ISOD and GE methods

Isodesmic (ISOD) reactions have been used as hypothetical reactions in thermochemistry, and are widely used in quantum mechanics for enthalpy of formation calculations. In the ISOD reaction, the types of chemical bonds formed in the products must be the same as the types of bonds broken in the reactants. The ISOD approach computes the  $\Delta H_f$  of the alkane, alcohol and ether molecules using the Eqs. 6–8, respectively.

$$\Delta H_f(C_nH_m) = H(C_nH_m) + a[H(CH_4) - \Delta H_f(CH_4)] - b[H(C_2H_4) - \Delta H_f(C_2H_4)] - c[H(C_2H_6) - \Delta H_f(C_2H_6)] \quad (6)$$

$$\Delta H_f(C_nH_mO_x) = H(C_nH_mO_x) + a[H(CH_4) - \Delta H_f(CH_4)] - b[H(C_2H_4) - \Delta H_f(C_2H_4)] - c[H(C_2H_6) - \Delta H_f(C_2H_6)] - d[H(CH_3OH) - \Delta H_f(CH_3OH)] \quad (7)$$

$$\Delta H_f(C_nH_mO_x) = H(C_nH_mO_x) + a[H(CH_4) - \Delta H_f(CH_4)] - b[H(C_2H_4) - \Delta H_f(C_2H_4)] - c[H(C_2H_6) - \Delta H_f(C_2H_6)] - e[H(CH_3OCH_3) - \Delta H_f(CH_3OCH_3)] \quad (8)$$

where  $H(C_nH_m)$  and  $H(C_nH_mO_x)$  are the calculated enthalpies of the molecules;  $H(CH_4)$  and  $\Delta H_f(CH_4)$  are the calculated enthalpies and experimental  $\Delta H_f$  of  $CH_4$ , respectively, and so on. The GE approach replaces the terms in brackets, such as  $[H(CH_4) - \Delta H_f(CH_4)]$  and so on, in the isodesmic reactions (Eqs. 6–8) with the group equivalents,  $\varepsilon_{CH_4}$ ,  $\varepsilon_{C_2H_4}$ ,  $\varepsilon_{C_2H_6}$ ,  $\varepsilon_{CH_3OH}$  and  $\varepsilon_{CH_3OCH_3}$ , respectively (as shown in Eqs. 9–11). These are thus least-squares fitted to reproduce the experimental  $\Delta H_f$  values of the training-set molecules. Table 1 summarizes the optimized group equivalents,  $\varepsilon_{CH_4}$ ,  $\varepsilon_{C_2H_4}$ ,  $\varepsilon_{C_2H_6}$ ,  $\varepsilon_{CH_3OH}$  and  $\varepsilon_{CH_3OCH_3}$ .

$$\Delta H_f(C_nH_m) = H(C_nH_m) + a\varepsilon_{CH_4} - b\varepsilon_{C_2H_4} - c\varepsilon_{C_2H_6} \quad (9)$$

$$\Delta H_f(C_nH_mO_x) = H(C_nH_mO_x) + a\varepsilon_{CH_4} - b\varepsilon_{C_2H_4} - c\varepsilon_{C_2H_6} - d\varepsilon_{CH_3OH} \quad (10)$$

$$\Delta H_f(C_nH_mO_x) = H(C_nH_mO_x) + a\varepsilon_{CH_4} - b\varepsilon_{C_2H_4} - c\varepsilon_{C_2H_6} - e\varepsilon_{CH_3OCH_3} \quad (11)$$

### CCAZD method

The correlation corrected atomization–dispersion (CCAZD) method was extended from the theoretical approach, CCAZ, developed in the previous studies for improving predictions of the  $\Delta H_f$  of carotenoids and xanthophylls [32]. The CCAZ method resembles the group equivalent methods, and is highly similar to Allinger's approach [20, 33]. The later approach adapts the procedure commonly used to calculate MM2/MM3/MM4 [34–36] values for  $\Delta H_f$ , and it works reasonably well for many kinds of compounds. The CCAZ approach corrects the  $\Delta H_f$  calculated from atomization scheme by adding group correction terms *I*,2-, *I*,3-, and *I*,4-corrections, as shown in Eq. 12. These correction terms correct the systematic errors in describing the proto-branching stabilization effects in DFT calculations, and are obtained by fitting the  $\Delta H_f$  of the training-set molecules to the experimental data using a least-squares fitting procedure:

$$\Delta H_f(C_nH_mO_x) = H(C_nH_mO_x) + \sum_i a_i E_i^{1,2} + \sum_j b_j E_j^{1,3} + \sum_k c_k E_k^{1,4} - n[H(C) - \Delta H_f(C)] - m[H(H) - \Delta H_f(H)] - x[H(O) - \Delta H_f(O)] \quad (12)$$

where  $a_i$  is the number of *i*th *I*,2-correction term  $E_i^{1,2}$  involved in the molecule, and so on. However, the CCAZD

method replaces the  $I,4$ -correction terms  $E_k^{1,4}$  in the Eq. 12 with an MM4-type dispersion function term  $E_{\text{DISP}}^{\geq 1,4}$  to model long-range dispersion energies more accurately, as shown by Eq. 13.

$$\begin{aligned} \Delta H_f(\text{C}_n\text{H}_m\text{O}_x) = & H(\text{C}_n\text{H}_m\text{O}_x) + \sum_i a_i E_i^{1,2} + \sum_j b_j E_j^{1,3} \\ & + \sum_{ii,jj} E_{\text{DISP}}^{\geq 1,4} \\ & - n[H(\text{C}) - \Delta H_f(\text{C})] - m[H(\text{H}) - \Delta H_f(\text{H})] \\ & - x[H(\text{O}) - \Delta H_f(\text{O})] \end{aligned} \quad (13)$$

and,

$$\sum_{ii,jj} E_{\text{DISP}}^{\geq 1,4} = \sum_{k,l}^{k \leq l} \left( \begin{array}{l} \text{type}(ii) = k; \\ \text{type}(jj) = l \\ \text{or} \\ \text{type}(ii) = l; \\ \text{type}(jj) = k \\ \sum_{ii < jj \text{ and } R_{ii,jj} \geq 0.85(r^k + r^l)} 184,000 \text{Exp} \left( \frac{-12R_{ii,jj}}{r^k + r^l} \right) - 2.25 \left( \frac{r^k + r^l}{R_{ii,jj}} \right)^6 \end{array} \right) E^{k,l} \quad (14)$$

where,  $k$  and  $l$  are the atom types of the interacting atoms  $ii$  and  $jj$ ;  $r^k$  and  $r^l$  are Bondi radii [37] of atom types  $k$  and  $l$ , respectively; Furthermore,  $R_{ii,jj}$  is the atomic distance between interacting atoms  $ii$  and  $jj$ ;  $E^{k,l}$  is a dispersion energy correction term for the interaction between atom types  $k$  and  $l$ . The empirical parameters  $E^{k,l}$ ,  $E^{1,2}$  and  $E^{1,3}$  (Scheme 1), were determined by fitting the calculated  $\Delta H_f$  of the training-set molecules to the experimental data through a least-squares fitting procedure. Note that MM4-type dispersion energies  $E_{\text{DISP}}^{\geq 1,4}$  were excluded when atomic distances  $R_{ii,jj}$  were shorter than 85% of sum of Bondi radii of those atoms ( $r^k + r^l$ ). The factor of 0.85 was chosen to prevent  $I,2$ - and  $I,3$ - corrections from being counted twice by the  $E_{\text{DISP}}^{\geq 1,4}$  term. Although this step function used in the CCAZD approach might lead to energy jump when atomic distances were around 85% of sum of Bondi radii, the results were generally acceptable since corrections were usually very small. Table 2 summarizes the optimized CCAZD correction parameters. The parameters are available along with the CCAZD program-

a written FORTRAN program to compute the CCAZD correction energy based on the provided structural information [38].

Isomerization energies, rotational barriers and conformational energies

To validate the performance of the CCAZD approach, 29 isomerization energies, 11 rotational barriers and 11 conformational energies were computed using B3LYP- and MPWB1K-CCAZD methods, and compared to the available experimental or reported data. The results predicted by MM4, G3, CBS-QB3, B2PLYP-D, B3LYP and MPWB1K were also obtained for comparison.

Potential surface of the cellobiose analog (CBA)

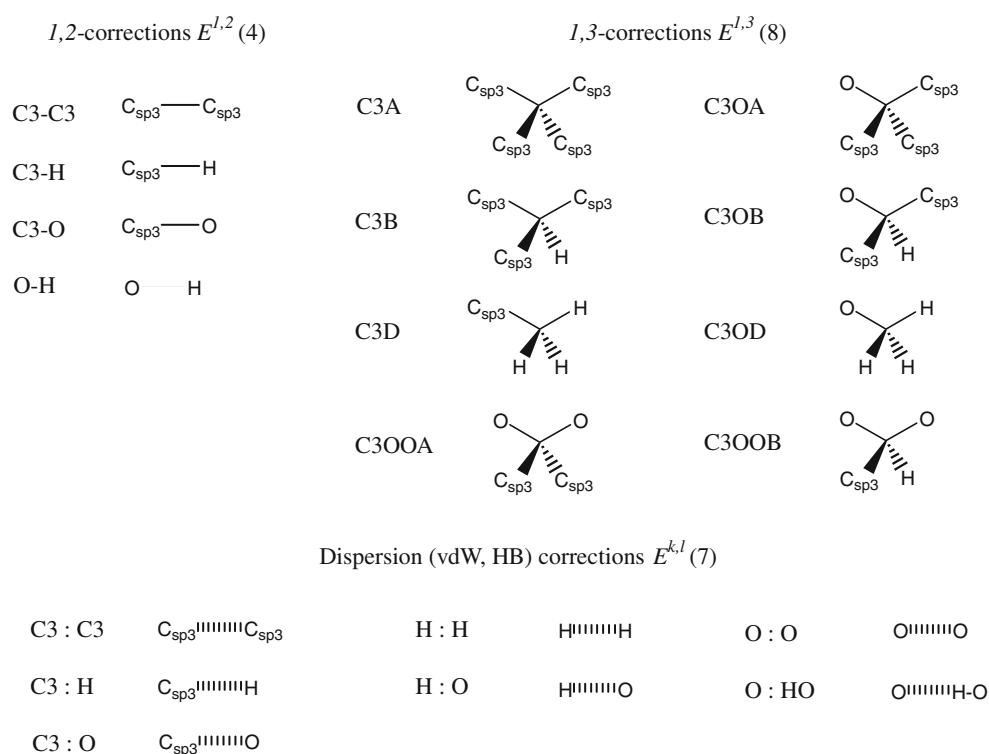
The potential surface of the cellobiose analog, comprising two *O*-linked tetrahydropyran (THP-*O*-THP) rings, was constructed by scanning the linkage torsion angles  $\varphi$  and  $\phi$  (Scheme 2), while optimizing all other bond, angle and torsion values. A step size of 20° increments for both  $\varphi$  and  $\phi$  was used over the range of 0° to 360°. Both B3LYP and B3LYP-CCAZD with 6-31G\* basis sets were used for constructing CBA potential surfaces.

## Results and discussion

### Enthalpies of formation

Tables 3 and 4 summarize enthalpies of formation ( $\Delta H_f$ ) of selected alkanes, alcohols and ethers from the training-set calculated by MM4, G3, CBS-QB3, B2PLYP-D, B3LYP (B3), MPWB1K (MP), B3LYP-AE (B3-AE), MPWB1K-AE (MP-AE), B3LYP-GE (B3-GE), MPWB1K-GE

**Scheme 1** Correlation correction terms used in the CCAZD scheme for alkanes, alcohols and ethers.  $C_{sp^3}$  represents  $sp^3$ -hybridized Carbon

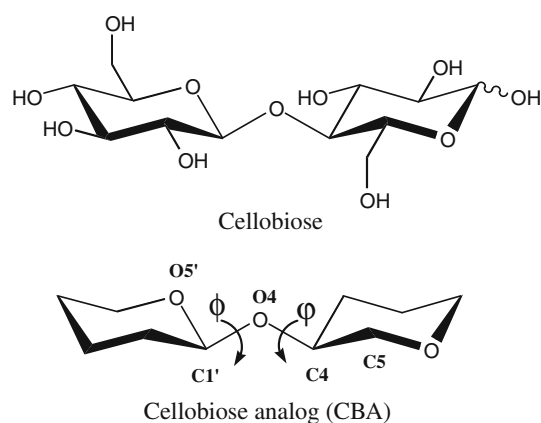


(MP-GE), B3LYP-CCAZD (B3-CCAZD) and MPWB1K-CCAZD (MP-CCAZD). As shown in Table 3, MM4 provides the best results for 52 alkanes with an average

absolute deviation (AAD) of 0.34 kcal/mol. B3- and MP-CCAZD follow with AADs of 0.45 and 0.50 kcal/mol, respectively. The G3 atomization (G3-AZ), CBS-QB3

**Table 2** The optimized correlation correction parameters used in the CCAZD scheme

	B3LYP/6-31G*		MPWB1K/6-31G*	
	In hartrees	In kcal/mol	In hartrees	In kcal/mol
C3-C3 (1,2-)	-0.0018402	-1.155	0.0035141	2.205
C-H (1,2-)	0.0004982	0.313	-0.0002552	-0.160
C3A (1,3-)	-0.0000602	-0.038	-0.0012657	-0.794
C3B (1,3-)	-0.0001690	-0.106	-0.0003992	-0.251
C3D (1,3-)	0.0000200	0.013	0.0000622	0.039
C3:C3 (DISP)	0.0000242	0.015	0.0000854	0.054
C3:H (DISP)	0.0003198	0.201	0.0000995	0.062
H:H (DISP)	-0.0001057	-0.066	0.0000739	0.046
C3-O (1,2-)	-0.0017566	-1.102	-0.0013866	-0.870
O-H (1,2-)	-0.0106288	-6.670	-0.0135815	-8.523
C3OA (1,3-)	0.0005590	0.351	-0.0001543	-0.097
C3OB (1,3-)	0.0006475	0.406	0.0006480	0.407
C3OD (1,3-)	0.0006285	0.394	-0.0001839	-0.115
C3OOA (1,3-)	0.0001395	0.088	0.0021254	1.334
C3OOB (1,3-)	-0.0001552	-0.097	0.0015316	0.961
C3:O (DISP)	0.0000838	0.053	0.0003043	0.191
H:O (DISP)	0.0000427	0.027	0.0001559	0.098
O:O (DISP)	-0.0018571	-1.165	-0.0013648	-0.856
O:HO (DISP)	-0.0014113	-0.886	-0.0021756	-1.365



**Scheme 2** Structures of cellobiose (CB) and cellobiose analog (CBA)

isodesmic (CBS-ISOD) and G3 isodesmic (G3-ISOD) results are also in good agreement with the experimental data with AADs of 0.54, 0.61 and 0.63 kcal/mol, respectively. The B3LYP and MPWB1K results of the AE and GE approaches are significantly better than those of the AZ and ISOD methods, though not as good as those of the CCAZD approach. In contrast, the results predicted by B2PLYP-D, with an AAD of 2.39 kcal/mol, are not as good as expected. The poor B2PLYP-D results (which were generally too negative) appeared in the highly branching and polycyclic molecules. This is likely because B2PLYP-D uses the same dispersion energy function for both inter- and intra- molecular interactions, which Schleyer et al. recommended should be separated. However, this problem does not appear in the CCAZD approach, since 1,2- and 1,3-correction terms handle the short-range corrections while an MM4-type dispersion function term handles long-range corrections.

There are three “problematic compounds” that deserve discussion: They are tri-*t*-butylmethane (**alkh24**), cyclododecane (**alkh31**) and dodecahedrane (**alkh56**). Both B3- and MP-CCAZD seemingly overestimate the  $\Delta H_f$  of these compounds, with deviations of 8.58, 1.50, 6.85 and 6.94, 1.54, 8.86 kcal/mol, respectively. For **alkh31**, the error might be on the experimental side [22]. Note that G3-AZ and G3-ISOD methods also overestimate the  $\Delta H_f$  of this compound by a similar amount (1–2 kcal/mol). The **alkh24** and **alkh56** values appear not to be very well calculated by CCAZD methods, with apparent errors of 7–9 kcal/mol. Compounds of this general type (globular hydrocarbons) tend to be extremely difficult to purify, however. Unfortunately, due to the size of the molecules, G3 and CBS-QB3 results are unavailable for compounds **alkh24** and **alkh56** for comparison.

As shown in Table 4, B3-CCAZD achieves the best results for 54 alcohol and ether compounds with an AAD

of 0.45 kcal/mol, followed by MM4, MP-CCAZD, CBS-ISOD, G3-ISOD, and G3-AZ, with AADs of 0.46, 0.52, 0.52, 0.68, and 0.73 kcal/mol, respectively. However, the results predicted by B2PLYP-D, with an AAD of 0.87 kcal/mol, are significantly better compared to those of alkanes. This might be due to relatively few highly branching and polycyclic molecules in the alcohol/ether training-set. Three “problematic compounds” are also present in the alcohol/ether calculations: They are 3-oxabicyclo[3.2.2]nonane (**alch30**) and 1,3-dioxacycloheptane (**alch41**). All methods, except B2PLYP-D, overestimated the  $\Delta H_f$  of **alch30** by  $\sim 2$  kcal/mol or more, and underestimated the  $\Delta H_f$  of **alch41** by  $\sim 3$  kcal/mol or more. Generally, the  $\Delta H_f$  of the training-set alkanes, alcohols and ethers predicted by B3- and MP-CCAZD agree well with the experimental data. The overall AADs for 106 training-set compounds are 0.45 and 0.51 kcal/mol, respectively.

Table 5 summarizes  $\Delta H_f$  of 20 test-set molecules including some 3- and 4-membered ring molecules. As shown in the table, the  $\Delta H_f$  of the test-set molecules predicted by B3- and MP-CCAZD with AADs of 1.45 and 1.64 kcal/mol, respectively, are very comparable to MM3, G3 and CBS-QB3 results, and they agree well with the experimental data [39] for most cases. The AADs for MM3, G3-AZ, CBS-ISOD and G3-ISOD results are 1.27, 1.56, 1.63 and 1.73 kcal/mol, respectively.

#### Isomerization energies, rotational barriers and conformational energies

The performance of an approach can be further validated by the accuracy of its predictions of energies of isomerization reactions. Table 6 summarizes 29 isomerization energies predicted by B3-CCAZD, MP-CCAZD, and other methods. This table shows that B3-CCAZD provides the best result for 22 isomerization energies with an AAD of 0.55 kcal/mol, followed by G3, CBS-QB3, MP-CCAZD, MM4 and B2PLYP-D results with AADs of 0.77, 0.77, 0.82, 1.16 and 1.23 kcal/mol, respectively. Table 6 shows that energies of isomerization reactions with the CCAZD treatment are improved significantly. The AADs for B3LYP and MPWB1K fall from 3.55 and 2.44 to 0.55 and 0.82 kcal/mol, respectively. However, both B3- and MP-CCAZD seemingly overestimate the energy of isomerization reaction tridecane  $\rightarrow$  tri-*t*-butylmethane by more than 5 kcal/mol. This discrepancy might be due to the experimental  $\Delta H_f$  of tri-*t*-butylmethane (**alkh24**) as indicated earlier. For the reactions 2-methyl-1,3-dioxane  $\rightarrow$  4-methyl-1,3-dioxane and 5,5-dimethyl-1,3-dioxane  $\rightarrow$  2-methoxytetrahydropyran, B3- and MP-CCAZD significantly underestimated their isomerization energies, as did G3, CBS-QB3, and B2PLYP-D.



**Table 3** Enthalpies of formation for selected training-set alkanes, shown as difference from the experimental values (kcal/mol)<sup>a</sup>

Label	Exp $\Delta H_f$	$\Delta MM4^c$	AZ (Atomization)					ISOD (Isodesmic)							
			$\Delta G3$	$\Delta CBS-$ QB3	$\Delta B3-$ AE <sup>c</sup>	$\Delta MP-$ AE <sup>c</sup>	$\Delta B3-$ CCAZD <sup>c</sup>	$\Delta MP-$ CCAZD <sup>c</sup>	$\Delta G3$	$\Delta CBS-$ QB3	$\Delta B2PLYP-$ D	$\Delta B3$	$\Delta MP$	$\Delta B3-$ GE <sup>c</sup>	$\Delta MP-$ GE <sup>c</sup>
Sign avg <sup>b</sup>		0.08	-0.07	2.42	10.92	-8.88	-0.01	0.10	0.14	0.16	-2.27	10.55	1.42	0.04	0.11
AAD <sup>b</sup>		0.34	0.54	2.42	11.04	8.90	1.67	1.03	0.45	0.50	2.39	10.55	5.20	1.46	0.70
Methane	-17.89	0.00	-0.16	0.16	-1.25	0.64	-0.06	-0.42	0.00	0.00	0.00	0.00	0.00	4.09	2.09
Ethane	-20.00	0.49	0.02	0.64	-1.28	-1.49	-1.56	-1.37	-0.27	-0.76	0.00	0.00	0.00	1.76	0.64
Propane	-24.82	-0.17	-0.25	0.67	-0.60	-3.17	-2.37	-1.90	-0.61	-0.96	-0.10	1.18	0.92	0.12	-0.39
Butane	-30.15	0.17	0.08	1.30	0.83	-4.09	-2.15	-1.37	-0.10	-0.39	0.31	2.88	2.37	-0.49	-0.36
Pentane	-35.00	-0.03	-0.05	1.53	1.89	-5.41	-2.35	-1.30	-0.05	-0.29	0.27	4.20	3.42	-1.52	-0.80
Hexane	-39.96	-0.16	-0.09	1.78	3.00	-6.67	-2.52	-1.21	0.02	-0.18	0.34	5.58	4.52	-2.52	-1.21
Heptane	-44.89	-0.27	-0.17	2.07	4.14	-7.92	-2.62	-1.03	0.17	0.01	0.38	6.99	5.65	-3.45	-1.53
Octane	-49.82	-0.38	-0.26	2.27	5.23	-9.21	-2.76	-0.92	0.26	0.14	0.41	8.35	6.72	-4.42	-1.93
Isobutane	-32.15	-0.21	0.18	1.39	2.06	-3.28	-1.18	-0.83	-0.46	-0.27	-0.07	4.11	3.18	0.48	0.18
Isopentane	-36.92	0.24	0.46	2.01	3.94	-4.05	-0.72	-0.37	-0.03	0.06	0.32	6.25	4.78	0.11	0.14
Neopentane	-40.27	-0.39	0.03	1.60	5.63	-2.97	0.91	0.65	-0.33	0.31	-1.03	7.95	5.86	1.74	1.15
Tri- <i>i</i> -butylmethane	-56.40	2.33			36.68	-0.25	20.13	12.73	8.58	6.94		41.13	27.51	14.33	9.22
Cyclopentane	-18.74	0.15	0.77	2.77	5.47	-6.85	-1.92	-0.99	0.63	0.42	-0.64	4.20	-3.02	0.36	0.39
Cyclohexane	-29.43	-0.16	0.60	2.38	6.45	-8.38	-2.42	-1.35	-0.14	-0.18	-0.93	5.44	-2.18	-0.97	-0.47
Cyclododecane	-54.59	1.06	1.34	5.56	19.60	-12.35	2.77	2.62	1.50	1.54	-0.71	20.19	8.05	-0.75	0.48
Norbornane	-13.12	0.03	0.00	1.85	12.28	-12.08	-0.73	-1.65	1.26	-0.34	-4.42	7.95	-8.52	1.34	-0.39
Adamantane	-31.76	-0.08	-1.56	1.77	20.30	-16.75	0.17	-0.57	0.35	-0.07	-7.83	13.18	-11.09	1.20	0.06
Dodecahedrane	18.20	-4.05			61.19	-33.93	4.88	11.83	6.85	8.86		27.99	-44.58	9.22	14.46

<sup>a</sup> The compounds listed in this table are part of 56 training-set alkanes. For complete results, please see Table 3S (supporting information). Experimental  $\Delta H_f$  are adopted from MM4 report [20]; B3 B3LYP, MP MPWB1K, B3-CCAZD B3LYP-CCAZD, MP-CCAZD MPWB1K-CCAZD, B3-AE B3LYP-AE, MP-AE MPWB1K-AE, B3-GE B3LYP-GE, MP-GE MPWB1K-GE

<sup>b</sup> The statistical results for the whole training-set. For comparison, entries with missing reported or any calculated data were excluded in statistic (4 excluded)

<sup>c</sup> Conformational population terms (POP, see supporting information Table 1S) were added to the calculated  $\Delta H_f$  of the most stable conformations to compare to the experimental  $\Delta H_f$

**Table 4** Enthalpies of formation for selected training-set alcohols and ethers, shown as difference from the experimental values (kcal/mol)<sup>a</sup>

Label	Exp $\Delta H_f$	$\Delta MM4^c$	AZ (Atomization)					ISOD (Isodesmic)									
			$\Delta G3$	$\Delta CBS-QB3$	$\Delta B3$	$\Delta MP$	$\Delta B3-AE^c$	$\Delta MP-AE^c$	$\Delta B3-CCAZD^c$	$\Delta MP-CCAZD^c$	$\Delta G3$	$\Delta CBS-QB3$	$\Delta B2P-LYP-D$	$\Delta B3$	$\Delta MP$	$\Delta B3-GE^c$	$\Delta MP-GE^c$
Sign avg <sup>b</sup>		0.12	-0.44	-0.70	7.88	3.54	0.50	0.41	0.08	0.03	-0.43	-0.17	-0.53	4.00	1.10	0.05	0.04
AAD <sup>b</sup>		0.46	0.73	1.27	7.88	4.52	3.71	2.89	0.45	0.52	0.68	0.52	0.87	4.04	3.33	1.05	0.58
Methanol	-48.16	-0.01	0.29	-0.46	6.24	10.09	3.11	2.45	-0.01	-0.03	-0.01	0.00	0.00	0.00	0.00	1.01	0.69
Ethanol	-56.24	0.42	0.31	-0.06	6.85	8.46	2.49	2.06	-0.07	-0.04	0.08	0.16	0.16	0.88	0.73	0.21	0.27
1-Propanol	-61.17	-0.17	0.33	0.26	7.85	6.89	2.38	1.87	0.01	-0.16	0.16	0.23	0.19	2.14	1.53	-0.08	-0.05
Isobutanol	-67.84	-0.16	0.05	0.32	9.83	5.95	2.90	1.96	-0.50	-0.44	-0.06	0.05	-0.37	4.39	2.96	0.26	-0.09
1-Butanol	-65.79	-0.02	0.25	0.57	8.89	5.74	2.41	2.18	0.22	0.21	0.14	0.31	0.33	3.45	2.74	-0.23	0.13
2-Butanol	-69.98	-0.28	-0.32	0.02	8.88	5.35	1.86	1.26	-0.16	-0.38	-0.43	-0.25	-0.44	3.44	2.35	-0.77	-0.79
t-Butanol	-74.72	0.23	-0.29	0.14	10.65	6.72	3.38	2.38	-0.24	-0.05	-0.40	-0.12	-0.84	5.21	3.73	0.74	0.33
1-Pentanol	-70.66	-0.08	0.07	0.69	9.88	4.36	2.36	2.25	0.34	0.37	0.02	0.19	0.27	4.71	3.73	-0.46	0.08
2-Me-2-Butanol	-79.06	0.10	-0.39	0.43	12.24	5.31	3.69	2.16	0.01	-0.40	-0.44	-0.08	-0.89	7.07	4.68	0.87	-0.01
1-Hexanol	-75.65	-0.02	0.05	1.02	11.07	3.17	2.53	2.52	0.65	0.72	0.06	0.27	0.37	6.17	4.91	-0.47	0.22
Cyclopentanol	-57.97	0.00	0.07	-0.01	12.11	1.75	5.24	4.29	-0.57	-0.23	-0.16	-0.93	-2.32	3.34	-3.87	-0.88	-0.54
Cyclohexanol	-69.31	-0.51	-0.16	0.89	14.13	0.89	5.90	4.56	-0.23	-0.24	-0.33	-0.26	-1.82	5.63	-2.36	-0.39	-0.40
3-Oxa-bicyclo- [3.3.2]nonane	-53.18	4.64	1.82	3.14	16.51	-6.00	6.82	5.15	1.78	1.65	1.93	1.93	-1.79	10.83	-5.08	3.67	2.57
1,3-Propane Diol	-97.61	-3.04	1.37	0.46	16.27	17.17	14.49	12.84	0.59	0.72	0.66	0.73	-0.75	0.46	-5.65	2.91	2.83
1,3-Dioxacycloheptane	-82.84	-0.41	-2.78	-3.95	4.15	-2.19	-2.64	-2.59	-2.66	-3.22	-2.79	-2.58	-3.50	-0.36	-8.39	-2.08	-2.21
Ethylene Glycol	-92.64	-0.12	-0.23	-1.58	13.19	16.14	8.68	6.59	-0.13	-0.13	-0.91	-0.74	-0.95	-0.27	-1.04	-0.52	-0.57
Water	-57.80	0.00	0.51	-0.23	14.06	18.14	12.30	9.38	0.72	1.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00

<sup>a</sup> The compounds listed in this table are part of 54 training-set alcohols and ethers. For complete results, please see Table 4S (supporting information). Experimental  $\Delta H_f$  are adopted from MM4 reports [42–44]. B3 B3LYP, MP MPWB1K, B3-CCAZD B3LYP-CCAZD, MP-CCAZD MPWB1K-CCAZD, B3-AE B3LYP-AE, MP-AE MPWB1K-AE, B3-GE B3LYP-GE, MP-GE MPWB1K-GE

<sup>b</sup> The statistical results for the whole training-set

<sup>c</sup> Conformational population terms (POP, see supporting information Table 2S) were added to the calculated  $\Delta H_f$  of the most stable conformations to compare to the experimental  $\Delta H_f$



**Table 5** Enthalpies of formation for test-set molecules, shown as difference from the experimental values (kcal/mol)<sup>a</sup>

Compounds	Exp $\Delta H_f^b$	$\Delta MM3$	AZ (Atomization)				ISOD (Isodesmic)	
			$\Delta G3$	$\Delta CBS-QB3$	$\Delta B3-CCAZD$	$\Delta MP-CCAZD$	$\Delta G3$	$\Delta CBS-QB3$
Sign avg		−0.07	0.84	1.13	0.85	1.05	1.00	1.12
AAD		1.27	1.56	2.32	1.45	1.64	1.73	1.63
Cyclopropane	12.74 (0.14)	0.21	0.92	1.60	1.31	0.33	1.10	0.88
1,1-Dimethylcyclopropane	−1.97 (0.28)	0.47	0.12	1.42	−1.39	−0.52	0.43	0.22
cis-1,2-Diethylcyclopropane	−10.66 (0.39)	−0.57	1.05	3.08	−0.66	0.09	1.48	1.39
trans-1,2-Diethylcyclopropane	−11.74 (0.43)	−0.86	0.65	2.66	−0.82	−0.16	1.08	0.97
Bicyclo[2.1.0]pentane	37.7	−0.40	0.83	1.78	−0.55	−0.81	1.05	0.49
Bicyclo[3.1.0]hexane	9.3 (0.8)	1.60	3.76	5.28	2.71	2.83	4.04	3.75
Bicyclo[4.1.0]heptane	0.4 (1.0)	0.10	3.61	5.55	2.19	2.52	3.95	3.78
1,1'-Bicyclopropyl	31.0 (1.2)	−0.44	2.39	3.85	1.89	1.80	2.66	2.32
Ethylene oxide	−12.58 (0.15)	−0.20	0.33	−0.61	0.64	1.47	0.34	0.36
Propylene oxide	−22.63 (0.15)	0.62	0.34	−0.19	0.43	1.61	0.41	0.54
Cyclohexene epoxide	−30.00 (0.26)	−1.04	0.54	1.00	0.76	1.58	0.70	0.92
Cycloheptene epoxide	−36.40 (0.74)	2.46	4.51	5.47	4.73	5.96	4.73	5.15
Oxetane	−19.25 (0.15)	−0.60	0.50	0.11	−0.41	0.38	0.57	0.84
Furfuryl alcohol	−88.2 (1.5)	−2.63	−2.55	−3.35	−2.12	−2.66	−2.80	−2.48
4-Methyl-1,3-dioxane	−90.45 (0.45)	−2.99	−1.28	−2.32	0.01	0.19	−1.20	−0.62
2-Methyl-1,3-dioxane	−95.00 (0.67)	−0.67	1.49	0.48	2.70	3.37	1.57	2.18
2,4-Dimethyl-1,3-dioxane	−102.3 (0.8)	−3.05	−1.61	−2.21	0.72	1.33	−1.47	−0.75
5,5-Dimethyl-1,3-dioxane	−100.50 (0.57)	1.77	2.98	2.24	3.46	3.47	3.12	3.70
1,3,6-Trioxocane	−111.70 (0.29)	4.57	−0.73	−3.23	0.87	−0.15	−0.76	−0.08
7-Norbornanol	−52.0 (2.0)	0.23	−0.96	0.02	0.57	−1.54	−1.08	−1.13

<sup>a</sup> B3-CCAZD B3LYP-CCAZD, MP-CCAZD MPWB1K-CCAZD<sup>b</sup> Experimental  $\Delta H_f$  are adopted from NIST reports [39]. The reported experimental errors are shown in parentheses, if any

Tables 7 and 8 summarize certain rotational barriers and conformational energies predicted by various methods. These tables show that all barriers and conformational energies predicted by B3- and MP-CCAZD remain almost unchanged after CCAZD treatment. However, the results predicted by other methods are slightly better by 0.10 kcal/mol in AADs.

## Carbohydrates

The CCAZD approach significantly improves the accuracy of predicting the enthalpies of formations and energies of isomerization reactions. However, a more challenging test is to apply the CCAZD approach to some bigger systems, such as carbohydrates. Due to the size of the systems, the conformational analysis of carbohydrates using quantum mechanics is always a challenge. Despite the computational cost-effectiveness of DFT, to use it to perform the conformational analysis of carbohydrates still suffers from its systematic error in describing proto-branching stabilization effects and long-range dispersion energy. However,

the DFT with empirical correction can hopefully improve the conformational analysis calculations on carbohydrates. Figure 1 presents the Ramachandran plots of the cellobiose analog (CBA) potential surfaces, calculated using B3LYP/6-31G\* (1A) and B3LYP/6-31G\*-CCAZD (1B). A single potential minimum appears on the B3LYP surface in the central region of interest in Fig. 1a, at  $\phi/\psi$  approximately  $-68/-136$ . The minimum energy area for CBA is more or less oval shaped and oriented north-south. There are also evident four large “mountains,” near the corners of the map. These are quite high, in the 11–16 kcal/mol range, according to DFT calculations. Because DFT calculations tend to significantly overestimate the energies in places where high van der Waals repulsions are present, these mountains should actually be smaller than indicated in Fig. 1a. Crystal structures are available for many cellobiose derivatives (disaccharides containing the CBA ring system). Thus, placing points on this contour map corresponding to the experimental X-ray structures of these CBA derivatives is possible [40, 41]. These points have been included on Fig. 1a, forming a

**Table 6** Energies of Isomerization Reactions ( $\Delta E_e$ , kcal/mol) for Selected Alkanes, Alcohols and Ethers

	Reported value <sup>a</sup>	MM4 <sup>b</sup>	G3	CBS-QB3	B2P LYP-D	B3	MP	B3-CCAZD	MP-CCAZD
<i>n</i> -Butane → Isobutane	−1.34	−1.55	−1.64	−1.59	−1.98	−0.48	−0.83	−1.80	−1.25
<i>n</i> -Pentane → Isopentane	−1.16	−0.65	−1.20	−1.12	−1.50	0.42	−0.24	−1.18	−0.82
<i>n</i> -Pentane → Neopentane	−3.99	−3.48	−4.50	−4.33	−5.56	−0.77	−1.98	−4.30	−3.32
<i>n</i> -Hexane → 2,2-Dimethylbutane	−3.08	−1.60	−3.49	−3.29	−4.53	1.13	−0.73	−3.01	−2.37
<i>n</i> -Hexane → 2,3-Dimethylbutane	−1.41	−0.52	−1.69	−1.52	−2.22	1.90	0.38	−1.31	−0.70
<i>n</i> -Heptane → 2,2,3-Trimethylbutane	−1.99	−0.41	−3.24	−2.92	−4.46	3.79	0.73	−2.26	−1.58
<i>n</i> -Heptane → 2,2-Dimethylpentane	−2.72	−1.51	−3.50	−3.28	−4.66	1.30	−0.58	−2.88	−2.29
<i>n</i> -Octane → Tetramethylbutane	−1.36	1.00	−3.82	−2.75	−5.03	7.53	2.69	−1.35	−0.70
<i>n</i> -Tridecane → tri- <i>t</i> -Butylmethane	21.65 <sup>c</sup>	29.42				44.65	34.14	28.85	27.23
<i>n</i> -Tetradecane → Octamethylhexane	23.72 <sup>c</sup>	30.98				46.74	30.92	27.43	22.60
<i>n</i> -Tetradecane → 3,3,4,4-Tetraethylhexane	18.83 <sup>c</sup>	26.98				35.32	22.64	21.32	16.11
Adamantane → <i>trans</i> -1-Methyl-1,2-dicyclopropylcyclopropane	78.08		80.89	80.85	79.49	75.28	78.82	77.65	79.28
1,3-Dimethyladamantane → 3-Methyl-tricyclo[4.4.1.0]undecane	22.8 <sup>d</sup>	21.28	22.81	22.28	23.68	18.78	18.99	21.63	20.35
1,3-Dimethyladamantane → 2,2-Dimethyladamantane	8.02	6.56	7.49	7.39	7.87	8.11	6.53	7.73	6.92
Methane Dimerization (tail to tail, D <sub>3d</sub> )	−0.51 <sup>e</sup>	−0.40	−0.38	−0.06	−0.32	0.20	−0.26	−0.38	−0.72
Ethanol → Dimethylether	12.35	12.03	11.74	11.54	12.16	5.50	7.06	12.44	12.35
<i>t</i> -Butanol → Diethylether	13.72	13.47	13.66	13.17	14.78	4.71	7.21	13.87	13.47
Cyclopentanol → T.H.P.	3.97	6.14	4.45	4.25	6.05	−2.61	−0.49	4.19	4.64
3,3-Dimethylpentane → T.H.P.	−19.42		−19.40	−19.60	−18.63	−20.51	−20.94	−18.18	−19.89
2-Methyl T.H.F. → T.H.P.		0.98	−0.37	−0.43	0.23	−0.66	−0.69	−0.98	−1.12
1-Methylcyclohexanol → <i>cis</i> -2-Methylcyclohexanol		1.78	1.88	1.69	1.99	0.66	1.07	1.36	1.55
1,2-Dimethoxyethane → 1,1-Dimethoxyethane		−10.76	−11.04	−10.77	−11.34	−9.60	−11.11	−11.66	−11.32
2-Methyl-1,3-dioxane → 4-Methyl-1,3-dioxane	4.79	3.66	1.62	1.65	1.24	2.05	1.97	1.77	1.27
5,5-Dimethyl-1,3-dioxane → 2-Methoxy T.H.P.	4.46	−0.93	0.63	0.65	1.79	−1.97	−1.17	0.84	0.60
1-Adamantanol → 2-Adamantanol	2.43	3.49	3.60	3.42	3.15	2.81	2.37	3.16	2.73
Water + Water → Water...Water (linear)	−5.44(70) <sup>f</sup> , −5.02 <sup>g</sup>	−4.59	−5.14	−5.07	−5.81	−7.68	−7.30	−5.20	−4.86
Water + Methanol → Water...Methanol (H-Donor) (linear)	−5.7 <sup>h</sup>	−4.39	−5.33	−5.17	−5.90	−7.73	−7.38	−5.55	−5.42
Methanol + Water → Methanol...Water (H-Donor) (linear)	−5.9 <sup>h</sup>	−5.10	−5.61	−5.23	−6.54	−7.51	−7.31	−5.37	−5.52
Methanol + Methanol → Methanol...Methanol (linear)	−5.7 <sup>i</sup>	−5.45	−5.85	−5.92	−6.74	−7.49	−7.33	−5.79	−6.09
AAD <sup>j</sup>		1.16	0.77	0.77	1.23	3.55	2.44	0.55	0.82

**Table 6** continued

	Reported value <sup>a</sup>	MM4 <sup>b</sup>	G3	CBS-QB3	B2P LYP-D	B3	MP	B3-CCAZD	MP-CCAZD
Outliers <sup>j,k</sup>		2	2	2	2	12	7	2	2

<sup>a</sup> Estimated from experimental  $\Delta H_f$  ( $\Delta E_e = \Delta\Delta H_f - \Delta H(\text{thermal}) - \Delta E(\text{Boltzmann})$ ), unless stated otherwise. Thermal enthalpies,  $H(\text{thermal})$ , used in the estimation were from the average of G3 and CBS-QB3 results. Energies contributed from Boltzmann population of conformers,  $E(\text{Boltzmann})$ , are significant for some cases, especially for those cases with long-chain involved, and they, except those of *n*-Tridecane and *n*-Tetradecane, were adopted from MM4 reports [20, 42–44].  $E(\text{Boltzmann})$  for *n*-Tridecane and etc. were computed using MM4 conformational search

<sup>b</sup> MM4 isomerization energies are calculated from MM4  $\Delta H_f$ ,  $H(\text{thermal})$  and  $E(\text{Boltzmann})$ .  $\Delta E_e = \Delta\Delta H_f - \Delta H(\text{thermal}) - \Delta E(\text{Boltzmann})$

<sup>c</sup> Thermal enthalpies,  $H(\text{thermal})$ , used in the estimation were from the average of B3 and MP results

<sup>d</sup> SCS-MP2/CBS(TZV2P – QZV3P) [19]

<sup>e</sup> CCSD(T)/aug-ccpVQZ complete base set limit [45]

<sup>f</sup> Estimated from experimental enthalpy of association of water dimer (–3.59 kcal/mol) [46]

<sup>g</sup> CCSD(T) limit [47]

<sup>h</sup> MP4/CBSB4 [48]

<sup>i</sup> High level ab initio calculation [48]

<sup>j</sup> For comparison, entries with missing reported or any calculated data were excluded in the statistic

<sup>k</sup> Calculated energies with opposite sign or absolute deviation is more than 3 kcal/mole were counted as outliers

pattern that is roughly centered on the oval shaped minimum, slightly to the west (left) and spilling over the 1.0 and 2.0 kcal contours still further to the west. The experimental pattern is not circular, but elliptical, with the major axis pointing roughly northeast-southwest. The corresponding B3LYP-CCAZD contour map (Fig. 1b) shows that the minimum energy area is wider (slightly to the south and west) than that in Fig. 1a. The four mountains are indeed in place, although slightly lower (in the range of 11–15 kcal/mol) than in Fig. 1a. Comparing Fig. 1b, a, the center of the data points is in approximately the same place.

Note the differences between B3LYP-CCAZD and B3LYP as shown by Ramachandran map (Fig. 2), where the points marked as “red triangle” and “red cross” are the potential minimum points for B3LYP-CCAZD and B3LYP contour maps, respectively. This figure shows that the CCAZD approach indeed lowers down the energies in the mountain regions, and flattens the minimum energy area.

## Conclusion

This study presents a new theoretical approach, called correlation corrected atomization–dispersion (CCAZD), to the empirical correction of density functional theory (DFT). The approach was improved from the CCAZ method developed in the previous studies for improving predictions of the enthalpies of formation ( $\Delta H_f$ ) of carotenoids and xanthophylls. In the CCAZD approach, 1,2- and 1,3-correction terms handle the corrections of the proto-branching stabilization problem of DFT, while an MM4-type of dispersion term handles the known long-range dispersion problem. In this study, the empirical corrections to the deficiencies in DFT were parameterized to reproduce reported  $\Delta H_f$  of training-sets containing alkanes, alcohols and ethers. The enthalpies of formation of the training-set compounds as calculated by CCAZD, combined with either two different DFT theories (B3LYP/6-31G\* and MPWB1K/6-31G\*) agreed well with the experimental data. The AADs for B3LYP- and MPWB1K-CCAZD results compared with experiments were 0.45 and 0.51 kcal/mol, respectively. The energies of isomerization reactions with the CCAZD treatment were also significantly improved. The AADs for 22 energies of isomerization reactions fell from 3.55 and 2.44 to 0.55 and 0.82 kcal/mol for B3LYP and MPWB1K, respectively. Finally, the CCAZD approach improves the cellobiose analog potential surface by flattening and shifting the minimum energy area in the direction where experimental data reside.

**Table 7** Rotational Barriers ( $\Delta E_c$ ) for Selected Alkanes, Alcohols and Ethers

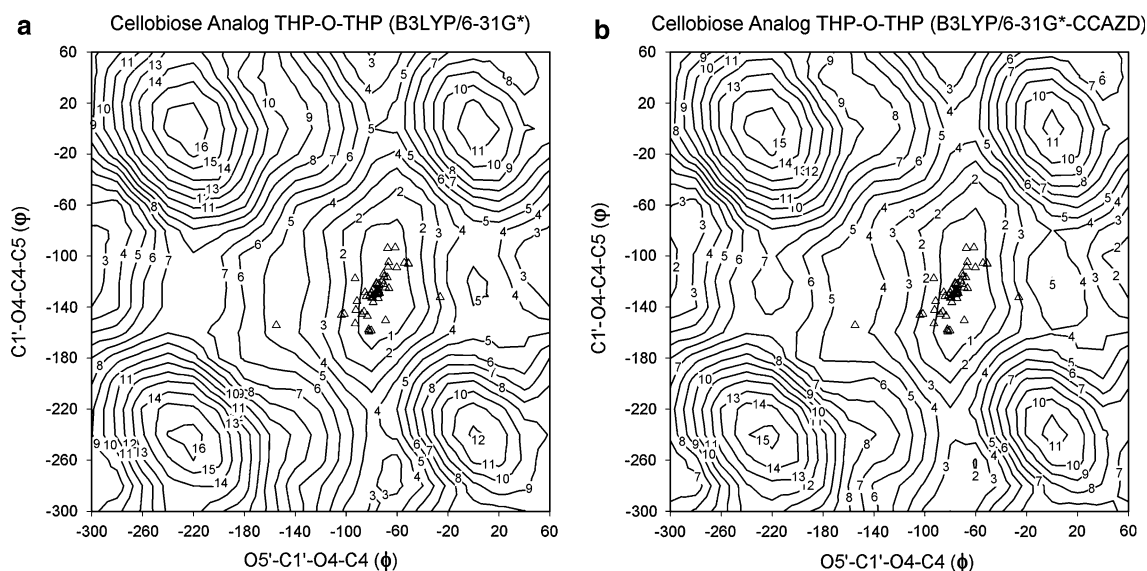
	Reported value	MM4	G3	CBS-QB3	B2P LYP-D	B3	MP	B3-CCAZD	MP-CCAZD
Ethane (H-C-C-H barrier)	2.88 <sup>a</sup>	2.60	2.86	2.86	3.15	2.82	2.91	2.77	2.95
Propane (C-C-C-H barrier)		2.81	3.17	3.16	3.34	3.10	3.20	3.24	3.31
Butane (C-C-C-C barrier)	5.40 <sup>b</sup>	4.94	5.62	5.58	6.10	5.69	5.40	5.12	5.27
Butane (C-C-C-H barrier)	3.33 <sup>b</sup>	3.40	3.40	3.40	3.38	3.35	3.43	3.65	3.56
Cyclohexane ( <i>Boat-Chair</i> )		7.20	7.57	7.57	7.82	7.34	7.23	7.15	7.31
Methanol (H-C-O-H barrier, <i>a</i> $\rightarrow$ <i>g</i> )	1.07 <sup>c</sup>	1.06	0.95	0.99	1.32	1.43	1.43	1.38	1.47
Ethanol (C-C-O-H barrier, <i>a</i> $\rightarrow$ <i>g</i> )	0.8–1.2 <sup>d</sup>	1.43	1.30	1.36	1.61	1.37	1.40	1.65	1.38
Ethanol (H-C-C-O barrier, <i>a</i> $\rightarrow$ <i>g</i> )	3.55 <sup>e</sup>	3.40	3.32	3.32	3.58	3.41	3.49	3.41	3.71
Propanol (C-C-C-O barrier, <i>g</i> $\rightarrow$ <i>a</i> )	5.43 <sup>f</sup>	5.09	5.23	5.15	5.77	5.15	5.08	5.09	5.17
Dimethoxymethane (C-O-C-O barrier, <i>a,g</i> $\rightarrow$ <i>a,a</i> )		2.73	5.62	5.58	6.34	5.31	5.08	4.99	4.89
1,2-Dimethoxyethane (O-C-C-O barrier, <i>a,a,a</i> $\rightarrow$ <i>a,g,a</i> )	8.78 <sup>g</sup>	8.42	8.71	8.55	9.22	8.42	8.56	8.00	7.97
AAD <sup>h</sup>		0.24	0.17	0.20	0.35	0.26	0.21	0.35	0.29
Outliers <sup>h,i</sup>		0	0	0	0	0	0	0	0

<sup>a</sup> Microwave spectroscopy [49]<sup>b</sup> High level ab initio calculations [50]<sup>c</sup> Millimeter spectroscopy [51]<sup>d</sup> Infrared spectroscopy [52]<sup>e</sup> Raman spectroscopy [53]<sup>f</sup> MP2/6-311 ++G(2d,2p) [42]<sup>g</sup> MP2/6-311 ++G(2d,2p) [44]<sup>h</sup> For comparison, entries with missing reported or any calculated data were excluded in the statistic<sup>i</sup> Calculated energies with opposite sign or absolute deviation is more than 3 kcal/mole were counted as outliers

**Table 8** Conformational Energies ( $\Delta E_e$ ) for Selected Alkanes, Alcohols and Ethers

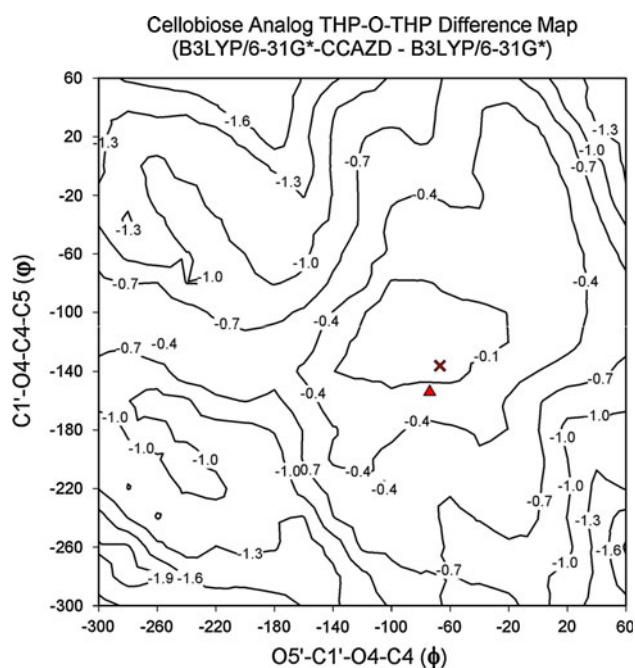
	Reported value	MM4	G3	CBS-QB3	B2P LYP-D	B3 <sup>k</sup>	MP	B3-CCAZD <sup>l</sup>	MP-CCAZD
Butane ( <i>a</i> → <i>g</i> )	0.61 <sup>a</sup>	0.68	0.59	0.62	0.60	0.85	0.61	0.45	0.46
Cyclohexane ( <i>chair</i> → <i>twist-boat</i> )	5.5 <sup>b</sup> , 6.18 <sup>c</sup>	6.27	6.13	6.33	6.07	6.54	6.03	6.29	5.98
<i>n</i> -C <sub>4</sub> H <sub>10</sub> folding (all <i>trans</i> → C6-C7-C8-C9 <i>gauche</i> )	2.2 <sup>d</sup>	4.93				5.94 (5.95)	4.65 (5.95)	3.79 (3.33)	3.23
<i>n</i> -C <sub>22</sub> H <sub>46</sub> folding (all <i>trans</i> → C10-C11-C12-C13 <i>gauche</i> )	−3.6 <sup>d</sup>	3.52				5.95 (9.10)	2.36 (5.95)	3.74 (−0.38)	−3.65
<i>n</i> -C <sub>30</sub> H <sub>62</sub> folding (all <i>trans</i> → C14-C15-C16-C17 <i>gauche</i> )	−8.8 <sup>d</sup>	1.44				5.98 (12.40)	0.12 (5.98)	3.75 (−3.88)	−10.19
Ethanol ( <i>a</i> → <i>g</i> )	0.40 <sup>e</sup> , 0.22 <sup>f</sup>	0.29	0.11	0.15	−0.13	−0.30	−0.17	−0.19	−0.19
Propanol ( <i>a</i> → <i>g</i> )	−0.30 <sup>g</sup> , −0.23 <sup>f</sup>	−0.18	−0.19	−0.19	−0.27	−0.14	−0.38	−0.16	−0.36
Dimethoxymethane ( <i>a,g</i> → <i>g,g</i> )	−2.89 <sup>h</sup>	−2.22	−2.77	−2.74	−2.78	−2.81	−2.98	−3.15	−3.18
1,2-Dimethoxyethane ( <i>a,a,a</i> → <i>a,g,a</i> )	0.31 <sup>i</sup>	0.03	0.32	0.32	0.31	0.52	0.68	1.05	0.84
Cyclohexanol ( <i>eq</i> → <i>ax</i> )	0.52 <sup>j</sup>	0.46	0.43	0.37	0.10	0.37	−0.16	0.36	−0.26
2-Methoxy T.H.P. ( <i>ax</i> → <i>eq</i> )	1.66 <sup>h</sup>	1.26	1.43	1.36	1.65	1.23	1.86	1.63	2.14
AAD <sup>m</sup>		0.25	0.13	0.17	0.15	0.32	0.27	0.30	0.37
Outliers <sup>m,n</sup>		0	0	0	1	1	2	1	2

<sup>a</sup> Estimated from experimental  $\Delta H$  (0.69 kcal/mol) [55]<sup>b</sup> Experiment [56]<sup>c</sup> MP2/aug-ccpVDZ [57]<sup>d</sup> MP2/aug-ccpVTZ//B3LYP-D/TZV(p,d) [14]<sup>e</sup> Experiment [54]<sup>f</sup> MP2/6-311 ++G(2d,2p) [42]<sup>g</sup> Estimated from experimental  $\Delta H$  (0.29 kcal/mole) [58]<sup>h</sup> MP2/6-311 ++G(2d,2p) [43]<sup>i</sup> MP2/6-311 ++G(2d,2p) [44]<sup>j</sup> Experiment [59]<sup>k</sup> Numbers in parentheses are computed with B3LYP/6-31G\*/MPWB1K/6-31G\*<sup>l</sup> Numbers in parentheses are computed with B3LYP/6-31G\*-CCAZD/MPWB1K/6-31G\*<sup>m</sup> For comparison, entries with missing reported or any calculated data were excluded in statistic<sup>n</sup> Calculated energies with opposite sign or absolute deviation is more than 3 kcal/mole were counted as outliers



**Fig. 1** Ramachandran plots of the CBA potential surfaces by B3LYP/6-31G\* (**a**) and B3LYP/6-31G\*-CCAZD (**b**) calculations. The numbers on the contours are energies relative to their minima at

about  $\phi = -68^\circ$ ,  $\phi = -136^\circ$  (**a**) and  $\phi = -74^\circ$ ,  $\phi = -154^\circ$  (**b**, estimated). Triangle on the contour maps are data points corresponding to the experimental X-ray structures of CBA derivatives



**Fig. 2** Ramachandran difference plot of the CBA potential surface (B3LYP-CCAZD-B3LYP). The points marked as red triangle and red cross are the potential minimum points for B3LYP-CCAZD and B3LYP contour maps, respectively

This study demonstrated that DFT calculations with CCAZD empirical corrections parameterized based on experimental enthalpies of formation achieved very good agreement with reported values for various chemical reactions, even with a basis set as small as 6-31G\*.

**Acknowledgments** The authors are indebted to Professor Norman L. Allinger for his advice and comments on this manuscript. The authors also thank the National Science Council of Taiwan, Republic of China, for support of this work, and the National Center for High-Performance Computing for computer time and facilities.

## References

1. Tirado-Rives J, Jorgensen WL (2008) J Chem Theor Comput 4:297–306
2. Wodrich MD, Wannere CS, Mo Y, Jarowski PD, Houk KN, Schleyer PvR (2007) Chem Eur J 13:7731–7744
3. Hobza P, Sponer J, Reschel T (1995) J Comput Chem 16: 1315–1325
4. Tsuzuki S, Luthi HP (2001) J Chem Phys 114:3949–3957
5. van MT, Gdanitz RJ (2002) J Chem Phys 116:9620–9623
6. Cerny J, Hobza P (2005) Phys Chem Chem Phys 7:1624–1626
7. Grimme S (2004) J Comput Chem 25:1463–1473
8. Elstner M, Hobza P, Frauenheim T, Suhai Sn, Kaxiras E (2001) J Chem Phys 114:5149–5155
9. Reha D, Kabeláč M, Ryjáček F, Šponer J, Šponer JE, Elstner M et al (2002) J Am Chem Soc 124:3366–3376
10. Wu Q, Yang W (2002) J Chem Phys 116:515–524
11. Zimmerli U, Parrinello M, Koumoutsakos P (2004) J Chem Phys 120:2693–2699
12. Grimme S (2006) J Comput Chem 27:1787–1799
13. Ducere J-M, Cavallo L (2007) J Phys Chem B 111:13124–13134
14. Schwabe T, Grimme S (2007) Phys Chem Chem Phys 9: 3397–3406
15. Chai J-D, Head-Gordon M (2008) Phys Chem Chem Phys 10: 6615–6620
16. Mackie ID, DiLabio GA (2008) J Phys Chem A 112: 10968–10976
17. Wodrich MD, Jana DF, Schleyer PvR, Corminboeuf Cm (2008) J Phys Chem A 112:11495–11500
18. Grimme S, Antony J, Ehrlich S, Krieg H (2010) J Chem Phys 132:154104/154101–154104/154119



19. Huenerbein R, Schirmer B, Moellmann J, Grimme S (2010) *Phys Chem Chem Phys* 12:6940–6948
20. Lii J-H, Allinger NL (2009) *J Mex Chem Soc* 53:96–107
21. Shamov GA, Budzelaar PHM, Schreckenbach G (2010) *J Chem Theor Comput* 6:477–490
22. Allinger NL, Chen K, Lii J-H (1996) *J Comput Chem* 17: 642–668
23. Hehre WJ, Ditchfield R, Radom L, Pople JA (1970) *J Am Chem Soc* 92:4796–4801
24. Curtiss LA, Raghavachari K, Redfern PC, Rassolov V, Pople JA (1998) *J Chem Phys* 109:7764–7776
25. Ochterski JW, Petersson GA, Montgomery JA Jr (1996) *J Chem Phys* 104:2598–2619
26. Grimme S (2006) *J Chem Phys* 124:034108/034101–034108/034116
27. Becke AD (1993) *J Chem Phys* 98:5648–5652
28. Lee C, Yang W, Parr RG (1988) *Phys Rev B* 37:785–789
29. Zhao Y, Truhlar DG (2004) *J Phys Chem A* 108:6908–6918
30. Baboul AG, Curtiss LA, Redfern PC, Raghavachari K (1999) *J Chem Phys* 110:7650–7657
31. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR et al (2003) G03. Revision B.05 edn. Gaussian, Inc., Pittsburgh
32. Lii J-H, Liao F-X, Hsieh H-Y, Hu C-H (2010) *J Phys Chem A* 114:12334–12344
33. Labanowski J, Schmitz L, Chen K-H, Allinger NL (1998) *J Comput Chem* 19:1421–1430
34. Allinger NL (1977) *J Am Chem Soc* 99:8127–8134
35. Allinger NL, Yuh YH, Lii J-H (1989) *J Am Chem Soc* 111: 8551–8566
36. Allinger NL, Chen K, Lii J-H (1996) *J Comput Chem* 17: 642–668
37. Bondi A (1964) *J Phys Chem* 68:441–451
38. CCAZD program reads Gaussian Formchk file (.fchk) and computes CCAZD energy correction for B3LYP/6-31G\* and MPWB1K/6-31G\* based on the given structural information. For the request of the program, please contact authors at jhrobert.lii@gmail.com or chingkh@cc.ncue.edu.tw
39. Afeefy HY, Liebman JF, Stein SE (2005) Neutral thermochemical data. In Linstrom PJ, Mallard WG (eds) NIST chemistry webbook, NIST standard reference database number 69, June 2005. National Institute of Standards and Technology, Gaithersburg, 20899. <http://webbook.nist.gov>
40. French AD, Johnson GP (2004) *Cellulose* (Dordrecht, Neth) 11:449–462
41. Lii J-H, Chen K-H, Johnson GP, French AD, Allinger NL (2005) *Carbohydr Res* 340:853–862
42. Allinger NL, Chen K-H, Lii J-H, Durkin KA (2003) *J Comput Chem* 24:1447–1472
43. Lii J-H, Chen K-H, Durkin KA, Allinger NL (2003) *J Comput Chem* 24:1473–1489
44. Lii J-H, Chen K-H, Grindley TB, Allinger NL (2003) *J Comput Chem* 24:1490–1503
45. Li AH-T, Chao SD (2009) *J Mol Struct THEOCHEM* 897:90–94
46. Curtiss LA, Frurip DJ, Blander M (1979) *J Chem Phys* 71:2703–2711
47. Kloppe W, van Duijneveldt-van de Rijdt JGCM, van Duijneveldt FB (2000) *Phys Chem Chem Phys* 2:2227–2234
48. Jursic BS (1999) *J Mol Struct THEOCHEM* 466:203–209
49. Riley KE, Hobza P (2007) *J Phys Chem A* 111:8257–8263
50. Hirota E, Endo Y, Saito S, Duncan JL (1981) *J Mol Spectrosc* 89:285–295
51. Allinger NL, Fermann JT, Allen WD, Schaefer HF III (1997) *J Chem Phys* 106:5143–5150
52. Lees RM, Baker JG (1968) *J Chem Phys* 48:5299–5318
53. Barrow GM (1952) *J Chem Phys* 20:1739–1744
54. Durig JR, Bucy WE, Wurrey CJ, Carreira LA (1975) *J Phys Chem* 79:988–993
55. Murphy WF, Fernandez-Sanchez JM, Raghavachari K (1991) *J Phys Chem* 95:1124–1139
56. Johnson WS, Bauer VJ, Margrave JL, Frisch MA, Dreger LH, Hubbard WN (1961) *J Am Chem Soc* 83:606–614
57. Kakhiani K, Lourderaj U, Hu W, Birney D, Hase WL (2009) *J Phys Chem A* 113:4570–4580
58. Abdurakhmanov AA, Ragimova RA, Imanov LM (1970) *Phys Lett A* 32:123–124
59. Hirsch JA (1967) In: Allinger NL, Eliel EL (eds) *Topics in stereochemistry*. Wiley Interscience, New York