# First-Principles Based Group Additivity Values for Thermochemical Properties of Substituted Aromatic Compounds

**Alper Ince, Hans-Heinrich Carstensen, Marie-Françoise Reyniers, and Guy B. Marin**
Laboratorium voor Chemische Technologie, Universiteit Gent, Technologiepark 914, B-9052 Zwijnaarde, Gent, Belgium

*DOI 10.1002/aic.15008*
*Published online August 28, 2015 in Wiley Online Library (wileyonlinelibrary.com)*

*A set of 7 Benson group additive values (GAV) together with 15 correction terms for non-nearest neighbor interactions (NNI) is developed to calculate the gas phase standard enthalpies of formation, entropies and heat capacities of monocyclic aromatic compounds containing methyl, ethyl, vinyl, formyl, hydroxyl, and methoxy substituents. These GAVs are obtained through least squares regression of a database of thermodynamic properties of 143 molecules, calculated at the post-Hartree–Fock G4 composite method. Out of the 15 NNIs, which account for several well-known substituent effects in aromatic molecules, 13 have been determined for the first time. All but two group additively calculated standard enthalpies of formation agree within 4 kJ mol$^{-1}$. The entropies and the heat capacities generally deviate less than 4 J mol$^{-1}$ K$^{-1}$ from the ab initio results. Natural bond orbital analysis is utilized to identify the underlying causes of the observed NNIs.* © 2015 American Institute of Chemical Engineers *AIChE J*, 61: 3858–3870, 2015
*Keywords: thermochemistry, group additivity, monocyclic aromatic hydrocarbons, lignin pyrolysis, non-nearest neighbor interactions*

## Introduction

The thermodynamic properties of molecules such as their standard enthalpy of formation ($\Delta_f H^\circ$), entropy ($S^\circ$), and heat capacity ($C_p$) play a pivotal role in the construction of fundamental kinetic models, which become increasingly more important in the optimization of industrial processes.[1–4] In recent years, such reaction networks have been used in various areas such as steam cracking of hydrocarbons,[4] combustion processes,[5,6] and the formation of polycyclic aromatic hydrocarbons.[7–9] Current interests in the thermochemical conversion of biomass to liquid fuels or commodity chemicals leads to the need for kinetic models that describe the pyrolysis of its components cellulose, hemicellulose, and lignin. The decomposition of lignin yields mainly substituted monocyclic aromatic compounds with substituent groups such as hydroxy (—OH), methoxy (—$OCH_3$), formyl (—CHO), vinyl (—C=$CH_2$), and alkyl (—R).[10–12] With few exceptions, reliable experimental thermochemical data for these species are not available. Furthermore, it is not feasible to use highly accurate first-principle calculations for each molecule encountered in such conversion processes because of the high demands in computing time and hardware resources. Therefore, to be able to construct a reliable kinetic model for lignin pyrolysis or other conversion processes that involve substituted monocyclic aromatic hydrocarbons, alternative methods play an important role.

A popular approach to describe thermodynamic properties of gas phase species is the group additivity (GA) method developed by Benson and coworkers.[13–16] Benson's basic idea is to divide molecules into smaller units, called groups, and use an additive scheme to obtain thermochemical data based on contributions from these groups, which are known as "group additive values (GAVs)." According to Benson, a group is defined as "a polyvalent atom in a molecule together with all its ligands."[16] Although GAVs account for the major part of a thermodynamic property, precise calculations require that additional contributions from interactions that extend beyond the range of a group have to be included as well. These so-called non-nearest neighbor interactions (NNIs) are incorporated into the GA method as correction terms. The thermodynamic property "$y$" thus is calculated as

$$y = \sum_{1}^{i} n_i * \mathrm{GAV}_i + \sum_{1}^{j} n_j * \mathrm{NNI}_j \qquad (1)$$

The $\mathrm{GAV}_i$ refer to different groups which occur $n_i$ times in the molecule, and $\mathrm{NNI}_j$ is the $j$th correction term which occurs $n_j$ times in the molecule. To use this method, all required GAVs and NNI correction terms must be known, and ongoing efforts aim to expand the GAV/NNI database and to improve the accuracy of existing entries.

Group values for Benson's GA method were initially derived entirely from experimental data,[15,17–20] but in the last

Additional Supporting Information may be found in the online version of this article.

Correspondence concerning this article should be addressed to M.-F. Reyniers at mariefrancoise.reyniers@ugent.be.

decade several studies used computational means to define these parameters.[21–28] It has been shown that GA methods are able to calculate standard enthalpies of formation with "chemical accuracy," meaning with a standard deviation of not more than 4 kJ mol$^{-1}$.[22–25] This accuracy is comparable to that of high level *ab initio* methods, which require substantially more time and effort. Benson's method has been successfully applied to standard enthalpies of formation of gas phase molecules,[4,15,17,22,24,25] and also to entropies and heat capacities.[13–16,23–25]

To ensure that the values obtained via GA are reliable, the GAV and NNI definitions have to be internally consistent. This means that (1) the GAV parameters are assigned in a systematic way resolving all linear dependencies,[14] (2) the set of NNIs are comprehensive and well defined, and (3) all parameters are determined from reference data of equal quality. Reyniers and coworkers[22–25] have started to create such an internally consistent GAV/NNI database based on *ab initio* results at the CBS-QB3 level of theory. Currently this database contains GAV and NNI definitions that allow calculation of thermodynamic properties for aliphatic hydrocarbons like alkanes, alkenes, alkynes, and few cyclic hydrocarbons and their corresponding radicals,[22,23] oxygenates such as alcohols, esters, ethers, ketones, carboxylic acids, ketenes and their corresponding radicals,[25] and organic sulfur compounds.[24] The purpose of this study is to extend the GAV and NNI database to substituted monocyclic aromatic hydrocarbons (MAHs). As this extension is motivated by the interest in modeling pyrolysis products of lignin, the following substituent groups are considered: hydroxy (—OH), methoxy (—$OCH_3$), formyl (—CHO), vinyl (—C=$CH_2$), and methyl (—$CH_3$) and ethyl (—$CH_2CH_3$). The following approach is adapted: (1) A database of accurate gas phase standard enthalpies of formation ($\Delta_f H^\circ$), entropies ($S^\circ$), and heat capacities ($C_p(T)$) for a set of 133 molecules is constructed using high level electronic structure calculations. (2) The necessary GAVs are identified and preliminary values are assigned based on the data for monosubstituted benzenes. (3) Deficiencies in the group additively calculated data are used to identify those non-nearest-neighbor interactions (NNIs) that contribute significantly to the thermodynamic properties. (4) Natural Bond Orbital (NBO) analyses are carried out to support the NNI selections. (5) All unknown GAV and NNI parameters are finally simultaneously optimized to best reproduce the thermodynamic data of the entire reference data set.

## Methodology

### *Electronic structure calculations*

The Gaussian 09 software package[29] has been used to perform all electronic structure calculations. Thermodynamic properties are calculated at the CBS-QB3[30] and G4[31] levels of theory. The CBS-QB3 composite method is widely used and has been shown to produce accurate enthalpies of formation[22,24,25] and rate coefficients.[32–35] It was the method of choice for deriving GAV and NNI parameters in the previous studies by Sabbe et al.,[22] Vandeputte et al.,[24] and Paraskevas et al.[25] and it would be the preferred method to extend the GAV database for aromatic species. However, CBS-QB3 calculated bond dissociation energies for the C—H bond in benzene (483.0 kJ mol$^{-1}$), the C—O bond in phenol (480.7 kJ mol$^{-1}$), and the aryl—CHO bond in benzaldehyde (428.2 kJ mol$^{-1}$) deviate more than 10 kJ mol$^{-1}$ from well-established experimental values (472.2 ± 2.2 kJ mol$^{-1}$,[36] 465.7 ± 4.2 kJ mol$^{-1}$,[36] and 415.5 ± 3.8 kJ mol$^{-1}$,[37] respectively). This casts doubts on the ability of the CBS-QB3 method to accurately describe aromatic molecules. For this reason, calculations were also performed at the G4 level of theory. G4 is also a composite method but differs from CBS-QB3 calculations in the details of the geometry optimization and the frequency analysis step as well as in the way the final energy is approximated by a series of single point energy calculations. The extrapolation methods to the basis set limit and the applied corrections to the final energy are also different. Details can be found in the original literature.[30,31] In the context of this study, it is worth noting that the G4 method yields extrapolated final electronic energies that are more accurate than the CBS-QB3 energies, however, at the cost of significantly higher CPU time demands, which effectively limits the applicability of G4 calculations to smaller molecules. With respect to the bond dissociation energies mentioned above, G4 yields those values to be 471.1 kJ mol$^{-1}$, 465.1 kJ mol$^{-1}$, and 414.2 kJ mol$^{-1}$, respectively, in excellent agreement with the experimental data.

### *Hindered rotor treatment*

To take thermochemical effects due to rotation of single bonds into consideration, the one-dimensional hindered rotor (1-D-HR) approach of Van Speybroeck and coworkers[38–40] is applied. It is well known that such a treatment significantly improves the calculated thermochemical properties,[39] in particular the entropy results. To calculate the contributions of internal rotations to the partition function and thermodynamic properties, the hindrance potentials for rotation are determined with relaxed potential energy surface scans at the B3LYP/6-31G(d) level and expanded to a truncated Fourier series using the first six sine and cosine terms. The internal rotations were identified via visual inspection of the animated vibrations. Only internal rotors with total barriers below 50 kJ mol$^{-1}$ are evaluated separately, all other modes are treated as a harmonic oscillators. Internal rotations with very low barriers (less than 1 kJ mol$^{-1}$) are considered free rotors. The reduced moment of inertia describing the rotation of the two molecule fragments around the axis which decouples internal from external rotation are calculated with the procedure described by Van Speybroeck et al.[39]

All thermochemical data calculated in this study are based on the lowest energy conformer of each molecule. The hindrance potentials of the internal rotors are used to identify low-energy conformers, which are then evaluated at the CBS-QB3 and G4 levels, respectively, to identify the lowest energy conformers. Note that this approach does not necessarily guarantee that the global minimum is found if internal rotors are coupled. The above described methodology is in line with the previous studies of Reyniers and coworkers.[22,24,25]

### *Standard enthalpy of formation*

The calculation of standard enthalpies of formation, which is the enthalpy relative to the composing elements in their standard state, always requires experimental information, as, for example, the enthalpy of carbon in its graphitic standard state cannot be determined accurately using *ab initio* methods. In the atomization energy method, the standard enthalpy of formation is obtained from the *ab initio* (AI) calculated enthalpy of atomization at 298 K and the experimental

**Table 1. BAC Values for CBS-QB3 and G4 Data**

| | BAC (kJ mol$^{-1}$) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | C—H | C—C | C=C | C—O | O—H | C=O | $C_b$—$C_b$ | $C_b$—H | $C_b$—O |
| CBS-QB3[a] | −0.19 | −2.07 | −3.45 | 1.58 | −1.77 | 3.11 | **0.73** | **−2.51** | **−0.98** |
| G4 | **−0.02** | **−1.09** | **0.00** | **−0.61** | **−1.44** | **0.50** | **2.04** | **−2.64** | **−4.77** |

[a]The nonaromatic BAC values for CBS-QB3 are taken from Paraskevas et al.[25]
The boldfaced values are derived in this study.

standard enthalpies of formation of atoms in the gas phase (gas,exp) at the same temperature, as given in Eq. 2

$$\Delta_f H^\circ\left(C_mH_nO_p; 298\ K\right) = m\Delta_f H^\circ_{gas,exp}(C) + n\Delta_f H^\circ_{gas,exp}(H) + p\ \Delta_f H^\circ_{gas,exp}(O) - [mH^\circ_{AI}(C) + nH^\circ_{AI}(H) + pH^\circ_{AI}(O) - H^\circ_{AI}(C_mH_nO_p)] \quad (2)$$

where $\Delta_f H^\circ_{gas,exp}(C) = 716.68$ kJ mol$^{-1}$, $\Delta_f H^\circ_{gas,exp}(H) = 218.0$ kJ mol$^{-1}$, and $\Delta_f H^\circ_{gas,exp}(O) = 249.18$ kJ mol$^{-1}$ at 298 K.[36]

The last four terms of Eq. 2. yield the *ab initio* (AI) calculated atomization enthalpies for $C_mH_nO_p$ at 298 K, that is, the reaction enthalpy of Eq. 3 at 298 K

$$C_mH_nO_p \rightarrow mC + nH + pO \quad (3)$$

Two different types of corrections have been applied to the atomization energies. First, spin-orbit (SO) corrections of 0.35 kJ mol$^{-1}$ per carbon atom and 0.93 kJ mol$^{-1}$ per oxygen atom[41] have been applied to the results from the CBS-QB3 calculations only since SO corrections are part of the G4 methodology.[31] Second, "Bond Additive Corrections" (BACs)[42] are applied to eliminate the remaining systematic errors of *ab initio* calculations

$$\Delta_f H^\circ(\text{BAC}) = \Delta_f H^\circ(\text{CBS-QB3 or G4}) + \sum_{ij} N_{ij}\text{BAC}_{ij} \quad (4)$$

where $i$ and $j$ run over the two different atom types and $N_{ij}$ is the number of bonds between atoms of type $i$ and $j$. As the CBS-QB3 and G4 methods suffer from different systematic bond-related errors, BAC values were developed separately for both methods.

In this study, BACs are used in accordance with the earlier studies,[22,25] however, since this work focuses on aromatic molecules, additional BAC values are defined to achieve optimal agreement between calculated and experimental enthalpies of formation. For this purpose, a training set of 77 molecules for which reliable experimental data is available, has been assembled. This training set, listed in Table S1 of the Supporting Information, contains 32 aliphatic hydrocarbons, 29 linear oxygenated hydrocarbons and 16 MAHs. The data for nonaromatic molecules were taken from the work of Paraskevas et al.[25] The rather small number of aromatic molecules present in this set can be explained with the scarcity of reliable experimental data for this class of molecules. The thermodynamic data for the aromatic molecules were retrieved from the "Thermochemical Database" of the NIST Webbook.[43] Several criteria were used to select the data. First, entries reporting $\Delta_f H^\circ$ values that were directly reported in the original study are preferred over those that have been deduced from published solid or liquid phase data. If the NIST Database contains several values for a given molecule, the entries with stated error margins are given priority. If several data with reported error bars exist, only those that overlapped with each other within the stated error margins were considered. Eventually, the average of the remaining data points was taken. Anisole (methoxybenzene) is the only molecule included in the training set that does not match the reliability criteria as there was no alternative methoxy (—$OCH_3$) substituted monoaromatic hydrocarbon available to be included in the BAC training set. For anisole, the average of the two available data points was taken even though they deviate by 8.8 kJ mol$^{-1}$ despite the fact that the reported error margins suggest accuracies of 1.2 kJ mol$^{-1}$[44] and 0.92 kJ mol$^{-1}$,[45] respectively.

Also provided in Table S1 of the Supporting Information are the deviations between CBS-QB3 (with SO corrections) and G4 calculated enthalpies of formation from the experimental data. It is evident, that in particular the CBS-QB3 results deviate substantially from the experimental values with the trend that the CBS-QB3 enthalpies are consistently higher than the measured ones. This is reflected in a mean deviation (MD) of 7.5 kJ mol$^{-1}$ for the entire training set. The results from G4 calculations are generally closer to the experimental values (MD = 3.1 kJ mol$^{-1}$), but G4 also tends to overestimate the enthalpies of formation.

To correct the theoretically calculated enthalpies of formation, a distinction between nine different bonds is made, which are C—C, C—H, C=C, C—O, O—H, C=O, $C_b$—H, $C_b$—$C_b$, and $C_b$—O. By differentiating between bonds of nonaromatic carbon atoms and those formed by an aromatic carbon atom, significant improvements of the corrected enthalpies of formation are achieved.

Out of the nine BACs for the bonds discussed above, six correction parameters have already been determined at the CBS-QB3 level of theory by Paraskevas et al.[25] Those were kept constant and only the remaining three parameters were optimized via unweighted linear regression. Since no previously defined BACs for G4 calculations were available, all nine parameters were subjected to the least squares fitting procedure. The BAC values for both levels of theory are reported in Table 1. The differences between bond additivity corrected enthalpies of formation and the experimental data are provided in Table S1 of the Supporting Information together with the statistics. The mean deviations (MD) between BAC-corrected and experimental $\Delta_f H^\circ$ data at 298 K are reduced to 0.0001 for both levels of theory and the mean absolute deviations (MAD) are 1.4 kJ mol$^{-1}$ (for both methods). These values are significantly improved compared to those of the uncorrected data (8.2 kJ mol$^{-1}$ and 3.3 kJ mol$^{-1}$, respectively). The maximum absolute deviations (MAX) are also clearly improved from 20.5 to 6.2 and 9.5 to 5.2 kJ mol$^{-1}$, respectively. The performance of BAC corrected CBS-QB3 and G4 calculations is further assessed using a test set of 14 MAHs. This test set contains MAHs for which the available experimental data do not fully meet the aforementioned criteria of reliability and thus they were not included in the training set. Given the higher uncertainty on the experimental values, the derived BACs are not expected to be as successful in improving the agreement between experimental and calculated enthalpies of

formation as for the training set. Nevertheless, BACs should lower the MAD for $\Delta_f H^\circ$ significantly at both levels of theory. Table S1 of the Supporting Information lists the test set molecules together with the experimental $\Delta_f H^\circ$ values. Also provided are the deviations of the theoretical enthalpies of formation (either CBS-QB3 or G4) from the experimental values prior to and after BAC are applied. As expected, the use of BAC clearly improves the agreement for this data set as well. For CBS-QB3, the MAD is reduced from 13.6 to 3.0 kJ mol$^{-1}$ whereas at the G4 level a reduction of the MAD from 5.9 to 3.4 kJ mol$^{-1}$ is observed.

A closer look at some data of the test set reveals that the experimental enthalpies of formation for *m*- and *p*-hydroxy phenol are not obtained accurately even when BACs are applied, while the $\Delta_f H^\circ$ value for o-hydroxy phenol (catechol), which surprisingly is higher than those for the *m*- and *p*-isomers despite the fact that a hydrogen bond can be formed, is well reproduced. As it is not obvious, why the calculations at both levels of theory would fail just for these two molecules, one might question the reliability of the experimental data. Certainly, a validation of these measurements would be helpful.

Another observation to be noted is that the bond additive correction values for G4 are very high for all three bonds that involve an aromatic carbon and that the $C_b$—$C_b$ and $C_b$—H BACs have similar magnitudes but opposite signs. Such large corrections appear to be in contrast to the good performance of G4 prior to the use of BAC. The $C_b$—H and $C_b$—$C_b$ values are strongly correlated because for the current data set, every $C_b$—H bond requires the presence of a $C_b$—$C_b$, and at the same time, a large fraction of $C_b$ atoms exist as $C_b$—(H) groups. The large BAC value found for G4 to correct errors in $C_b$—O bonds is expected to have a high uncertainty, because only a very limited number of enthalpies of formation were available to obtain this parameter and some stated inaccuracies are rather high (e.g. anisole, as discussed before).

As discussed in earlier sections, some CBS-QB3 calculated bond dissociation energies deviate from established experimental values, while the corresponding G4 results agree well with the measurements. Given that rather high BAC values for the G4 method were found, it is important to verify that the bond dissociation energies still agree with experiment. For the C—H bond in benzene, the C—O bond in phenol and the Aryl—CHO bond in benzaldehyde, the G4$^{BAC}$ bond dissociation enthalpies are 474.7 kJ mol$^{-1}$ (vs. 472.2 kJ mol$^{-1}$), 469.9 kJ mol$^{-1}$ (vs. 465.7 kJ mol$^{-1}$), and 415.3 kJ mol$^{-1}$ (vs. 415.5 kJ mol$^{-1}$), respectively. BAC-corrected CBS-QB3 values (485.4, 481.7, and 428.1 kJ mol$^{-1}$) still differ considerably more from the experimental values.

The similarly good performances of the CBS-QB3/BAC and G4/BAC calculation methods with respect to the enthalpies of formation of closed-shell aromatic molecules in the training and test sets (see Supporting Information Table S1) indicate that the poor bond dissociation enthalpy data obtained with CBS-QB3 are related to a problem with the enthalpy of formation for the phenyl radical.

### *Standard intrinsic entropies*

The entropy for a molecule calculated via statistical methods does not take symmetry or contributions from steric centers into account. Therefore, the initial entropies need to be corrected for such contributions if the total entropy of a molecule is needed. Conversely, GAV and NNI parameters in Benson's GA method[14] are "symmetry-free," because they relate to moieties of a molecule and not to the entire molecule. Since the goal of the current study is to determine such GAVs and NNIs, the symmetry-independent "intrinsic" entropies $S^o_{int}$ are used

$$S_{int}{}^\circ = S^\circ + R\ \ln\left(\frac{\sigma}{n_{opt}}\right) \tag{5}$$

where $\sigma$ is the global symmetry number and $n_{opt}$ is the number of optical isomers. The global symmetry number $\sigma$ is the product of external symmetry number $\sigma_{ext}$ and the internal symmetry numbers $\sigma_{int}$

$$\sigma = \sigma_{ext} \prod_k \sigma_{int,k} \tag{6}$$

### *Derivation of GAV and NNI values*

The GAV and the NNI values are determined by optimizing the agreement between group additively calculated data and a large set of thermodynamic values for molecules that contain these groups. The latter data is obtained via electronic structure calculations according to the methodology discussed in the previous paragraphs. The following overall strategy has been followed:

1. BAC corrected CBS-QB3 and G4 results are shown to compare equally well to experimental data. In anticipation of follow-up studies on radicals, the G4 method was chosen as reference for GAV development.
2. Preliminary GAVs are assigned using results for benzene and the six monosubstituted benzenes. The initial assignment of GAVs is needed to aid the identification of NNIs in the following step.
3. Application of the preliminary GAVs to all 63 double substituted benzenes reveals the need of NNIs and allows identification of those.
4. The GAVs and NNIs are simultaneously optimized using the full set of training molecules (133).
5. A test set consisting of 10 smaller MAH and 15 large MAHs is assembled to validate the transferability of the GAV/NNI data.
6. Finally, the 10 G4 calculated entries of the test set are added to the training set and a final optimization yields the best set of GAV/NNIs.

Only the final GAV and NNI parameters are reported in the manuscript whereas preliminary and intermediate values of GAVs and NNIs are documented in the Supporting Information.

Parameter optimization relied on the unweighted least-squares procedure. In this linear regression analysis, the following objective function is minimized

$$\mathrm{SSQ} = \sum_i^n (y_i - \hat{y}_i)^2 \tag{7}$$

In Eq. 7, $y_i$ is the *ab initio* calculated enthalpy of formation ($\Delta_f H^\circ$), entropy ($S^\circ$) or heat capacity ($C_p$(T)) of the molecule $i$, and $\hat{y}_i$ is the calculated thermochemical value via GA. This results in the equation

$$\overline{\mathrm{GAV/NNI}} = \left(X^T X\right)^{-1} X^T \bar{y} \tag{8}$$

In Eq. 8, $\overline{\mathrm{GAV/NNI}}$ is the calculated vector of the GAVs and NNIs parameters and $X$ is the occurrence matrix where in each row, there is a molecule and in each column, there is

Table 2. GAVs Taken from Earlier Studies by Reyniers and Coworkers[22,23,25]

| Fixed GAVs | GAV taken from | $\Delta_f H^\circ$ (kJ mol$^{-1}$) | $S^\circ$ (J mol$^{-1}$ K$^{-1}$) | $C_p$ (J mol$^{-1}$ K$^{-1}$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 300 K | 400 K | 500 K | 600 K | 800 K | 1000 K | 1500 K |
| O—($C_b$)(H) | O—($C_d$)(H)[25] | −188.1 | 106.3 | 24.6 | 30.3 | 32.5 | 33.2 | 33.3 | 33.6 | 35.0 |
| CO—($C_b$)(H) | CO—($C_d$)(H)[25] | −128.3 | 129.3 | 27.3 | 34.0 | 39.4 | 43.7 | 50.2 | 54.2 | 60.8 |
| $C_d$—($C_b$)(H) | $C_b$—($C_d$)(H)[22,23] | 30.4 | 25.7 | 18.1 | 24.1 | 29.1 | 32.7 | 37.1 | 40.0 | 43.1 |
| C—($C_b$)($H)_3$ | C—($C_d$)($H)_3$[22,23] | −42.9 | 127.2 | 25.0 | 31.8 | 38.2 | 43.9 | 53.2 | 60.5 | 72.3 |
| C—(O)($H)_3$ | From Ref. 25 | −42.9 | 127.1 | 25.3 | 32.1 | 38.4 | 44.1 | 53.4 | 60.6 | 72.5 |
| $C_d$—($H)_2$ | From Refs. 22,23 | 25.1 | 115.8 | 20.6 | 25.8 | 30.8 | 34.9 | 41.4 | 46.4 | 54.6 |
| C—(C)($H)_3$ | From Refs. 22,23 | −42.9 | 127.1 | 25.3 | 32.1 | 38.4 | 44.1 | 53.4 | 60.6 | 72.5 |

either a GAV or NNI. The elements of this matrix $X_{i,j}$ specify the number of occurrences of group $j$ in molecule $i$. Here, $\overline{y}$ is the vector which contains the *ab initio* thermodynamic property of interest. If a group always appears in connection with another group, they are linearly dependent and the values of both groups cannot be determined independently. To eliminate such linear dependencies, several GAV values have to be assigned arbitrarily. In this study, this assignment is based on structurally similar groups, as will be discussed later in more detail.

Seven GAVs found in the molecules of the database have already been defined in an earlier study by Reyniers and coworkers.[22,23,25] These values were adopted and not subjected to the optimization procedure.

A statistical analysis is performed to assess the reliability of the linear regression step and the quality of the optimization procedure. The quality of the fits is expressed in terms of the significance (F) of the regression, the MAD, root mean square (RMS) deviation, and maximum deviation (MAX) between *ab initio* data and data obtained via GA. The reported significance $F$ of the regression is calculated with the following equation

$$F = \frac{\sum_{i=1}^{n} \frac{\hat{y}^2}{p}}{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n-p}} \tag{9}$$

$n$ is the number of molecules and $p$ is the total number of GAV and NNIs.

The NBO method of Weinhold and coworkers has been employed to interpret the underlying physical origin of the NNIs.[46] The NBO analysis 6.0 program[47] transforms the many electron molecular wavefunction obtained from the Gaussian calculation into localized orbitals that are termed as "Natural Atomic Orbitals (NAOs)." The program delivers populations of core, valence and higher energy Rydberg orbitals of every single atom as a part of Natural Population Analysis. Optimized linear combinations of NAOs form natural hybrid orbitals (NHOs). Combination of NHOs leads to bonding and antibonding NBOs, which are localized orbitals that describe the Lewis-like molecular bonding pattern of electron pairs. From the idealized Lewis picture point of view, all filled NBOs are expected to have an occupancy of 2 and any departure from this idealized localized picture is interpreted as the presence of delocalization. The NBO calculations are carried out with results from single point energy calculation at the CCSD(T)/6-31G(d′) level at the B3LYP//CBSB7 optimized geometry.

## Results and Discussion

CBS-QB3/BAC and G4/BAC calculations have been performed to construct databases of accurate gas phase standard enthalpies of formation ($\Delta_f H^\circ$), entropies ($S^\circ$), and heat capacities ($C_p(T)$, with T = 300 K, 400 K, 500 K, 600 K, 800 K, 1000 K, and 1500 K) for 133 MAHs. The sets include benzene, six single substituted benzenes, all possible 63 double substituted benzenes, and 63 triple substituted benzenes.

A comparison of the BAC-corrected standard enthalpies of formation obtained at both levels of theory reveals that both datasets agree well with each other. The MAD is 1.19 kJ mol$^{-1}$ and only a few molecules out of 133 $\Delta_f H^\circ$ values deviate more than 4 kJ mol$^{-1}$. MAD for intrinsic entropies ($S^o_{int}$) is 0.63 J mol$^{-1}$ K$^{-1}$ and that for the heat capacities ($C_p$) at 300 K is 0.4 J mol$^{-1}$ K$^{-1}$. Histograms showing the distribution of these deviations are presented in **Figures S2** and **S3** of Supporting Information for standard enthalpies of formation and entropies, respectively. Based on the good agreement, either database is suitable to determine the GA parameters (GAV, NNI) for the MAHs of interest. Since the G4 method is more reliable for radicals from MAHs than CBS-QB3, the G4 database will be utilized as reference thermochemical database in this study.

### *Assignment of initial GAVs*

In total 14 GAVs are required for the GA calculation of the thermodynamic properties of the 133 MAHs of the reference dataset. Among these, there are 3, namely C—(O)($H)_3$, C—(C)($H)_3$, and Cd—($H)_2$, that describe nonaromatic moieties. These are taken from earlier studies[22,23,25] to ensure internal consistency with previous work. Due to linear dependencies, four additional GAVs cannot be determined independently from the remaining GAVs: O—($C_b$)(H), CO—($C_b$)(H), $C_d$—($C_b$)(H), and C—($C_b$)($H)_3$. The O—($C_b$)(H) always appears together with a $C_b$—(O) group, a CO—($C_b$)(H) group can only exist if also a $C_b$—(CO) group is present, the $C_d$—($C_b$)(H) is always accompanied by a $C_b$—($C_d$) group and the C—($C_b$)($H)_3$ requires the presence of the $C_b$—(C) group. The values for these four GAVs are set equal to the structurally similar GAVs O—($C_d$)(H), CO—($C_d$)(H), $C_d$—($C_d$)(H) and C—($C_d$)($H)_3$, respectively, as determined in the earlier studies by Sabbe et al.[22] and Paraskevas et al.[25] All predefined GAVs are listed in Table 2. Once the linear dependencies in the occurrence matrix $X$ are eliminated, the thermodynamic data of benzene and the six monosubstituted benzenes are used to provide initial assignments of the remaining independent GAVs. Table S4 in Supporting Information lists these initial GAVs for standard enthalpies of formation ($\Delta_f H^\circ$), entropies ($S^\circ$), and heat capacities ($C_p$).

The preliminary GAVs have been used to obtain the standard enthalpies of formation ($\Delta_f H^\circ$) and entropies ($S^\circ$) of all 63 double substituted benzenes in the reference database. Table S5 in Supporting Information lists those double substituted MAHs for which significant deviations between the

**Table 3. NNIs That Are Identified from the Differences Given in Table S5 in the Supporting Information**

| NNI | Interacting substituents | Molecules used in identifying NNIs | Contributions to NNI |
|---|---|---|---|
| **NNI1** | *o*-OH+CHO | 29 | HB[a], Mesomeric, Inductive |
| **NNI2** | *o*-CHO+CHO | 14 | HB, Mesomeric, Inductive |
| **NNI3** | *o*-MeO+MeO | 11 | Anomeric, Mesomeric, Inductive |
| **NNI4** | *o*-CH=$CH_2$+CHO | 53 | Steric, Inductive |
| **NNI5** | *o*-CH=$CH_2$+CH=$CH_2$ | 17 | Steric, Inductive |
| **NNI6** | *p*-OH/MeO+OH/MeO | 10,13,28 | Mesomeric, Inductive |
| **NNI7** | *p*-CHO+CHO | 16 | Mesomeric, Inductive |
| **NNI8** | *o*-Me/Et+CHO | 56,59 | Steric, Mesomeric, Inductive |
| **NNI9** | *o*-CH=$CH_2$+Me/Et | 62,65 | Steric |
| **NNI10** | *m*-CHO+CHO | 15 | Inductive |
| **NNI11** | *p*-CHO+OH/MeO | 31,43 | Mesomeric, Inductive |
| **NNI12** | *o*-MeO+CHO | 41 | HB, Mesomeric, Inductive |
| **NNI13** | *o*-Me/Et+Me/Et | 20,23,68 | Steric |
| **NNI14** | *o*-OH+OH/MeO | 8,26 | HB, Mesomeric, Inductive |
| **NNI15** | *o*-CH=$CH_2$+OH/MeO | 32,44 | Steric, Inductive |

The full list of molecules and their molecule numbers are given in Figure S1 of the Supporting Information.
[a]HB: Hydrogen Bond.

reference data and the GAV-estimates can be found. Some deviations exceed 20 kJ $mol^{-1}$ for the standard enthalpy of formation ($\Delta_f H^\circ$) and a large entropy difference of 23.1 J $mol^{-1}$ $K^{-1}$ is seen in one case. This clearly shows that NNI need to be accounted for to obtain reliable data from GA calculations. It is noticed that the largest deviations are found for substituents in *ortho* position, but several *para* substituted MAHs also display large deviations between the reference and group additively obtained data using the preliminary GAVs only. The observed differences in Supporting Information Table S5 are used in the next section to identify suitable NNIs and to explain the physical origin of these interactions.

*Identification of NNIs.* Since the database of molecules includes all possible 63 double substituted MAHs that can be obtained with hydroxy (—OH), methoxy (—$OCH_3$), formyl (—CHO), vinyl (—CH=$CH_2$), methyl (—$CH_3$), and ethyl (—$CH_2CH_3$) groups, all the possible binary interactions between these substituents are captured. The differences between *ab initio* and group additively obtained data for double substituted MAHs (see Supporting Information Table S5) have been used to identify the most important non-nearest neighboring interactions (NNIs). In Table 3, these NNIs are listed together with (a) the molecules in which they occur, and (b) the possible origin of these interactions. Initial NNI values for $\Delta_f H^\circ$, $S^\circ$, and $C_p$ (300 K) are given in Table S6 of the Supporting Information.

The general types of interactions between substituents of benzene are well established. The interactions given in Table S6 of the Supporting Information include hydrogen bonds (HBs), mesomeric, inductive, steric, and anomeric effects. Some of these interactions may occur alone, others occur always in combinations. The use of NBO analysis, which reveals information about charge transfers between different NBOs, is useful to develop an understanding of the intramolecular interactions.

Among the 15 NNIs identified, 4 NNIs contain significant contributions from HB to the overall intramolecular interaction. They are listed in Table S7 of the Supporting Information together with the stabilization energies ($E(2)_{HB}$) due to charge transfer between the donor lone pair NBO of an oxygen atom and the acceptor O—H or C—H Bond NBOs calculated by Second-Order Perturbation Theory Analysis as described in Ref. 46. Table S7 of the Supporting Information demonstrates that stronger HBs correlate with shorter HB distances and also shows that hydrogen bond formation appears to be the major contributor to **NNI1**, while it plays a minor role in the overall destabilizing **NNI2** and **NNI12**.

Mesomeric effects are investigated via the charge distributions obtained from the NBO analysis. The natural charges on five arene carbons of benzene and the ortho, meta, and para positions in single substituted MAHs are given in Table S8 of the Supporting Information. Also provided is the sum of these charges, which can be considered as a qualitative measure of the amount of charge transfer via resonance. Note, that the ring carbon which is substituted, is not included, because its charge results not only from resonance but also from inductive charge transfer. In benzene, the natural charge is equally distributed among all six carbons and the negative value of −0.235, which is counterbalanced by the positive charges on the hydrogens, indicates that aryl carbon atoms are more electronegative than hydrogen. An alkyl substituent (—$CH_3$/—$CH_2CH_3$) does not significantly alter the natural charge distribution or the total charge. The slight increase of the natural charges on the *ortho* and *para* carbon atoms indicates a very weak electron donating effect. In styrene ($C_6H_5$—CH=$CH_2$), the total charge of the five nonsubstituted carbon atoms is only slightly lower than in benzene indicating a small electron withdrawing effect of the —CH=$CH_2$ group. Hydroxy (—OH), and methoxy (—$OCH_3$) substituents increase the electron density on the ring carbons in *ortho* and *para* position. At the same time, the electron density on the ring carbons in *meta* position is reduced. The total charge is increased, which is consistent with the +M effect assigned to these substituents. In benzaldehyde ($C_6H_5$—CHO), a reverse trend is noticeable: the presence of the formyl (—CHO) group withdraws electrons from the ring carbons in *ortho* and *para* position, while a small charge increase is observed in the *meta* position. The total charge of the five nonsubstituted carbon atoms is lower than in benzene due to the dominant −M effect.

The above discussion is qualitatively supported by results from second-order perturbation theory analysis. A strong mesomeric charge transfer (+M effect) from the lone pair NBO on oxygen atom (donor) to the closest $\pi$-bond NBO on the arene ring (acceptor) is found in anisole and phenol. The mesomeric effect in benzaldehyde is in the opposite direction (−M): charge is transferred from one of the $\pi$-bond NBOs on the arene ring to the $\pi$-bond NBO of the double bond of the —CHO group. In styrene, charge transfer is observed in two

directions: from arene ring to the vinyl (—CH=$CH_2$) group and from the vinyl group to the arene ring. Both transfers effectively cancel out.

In double substituted MAHs, the total mesomeric effect is not simply the sum of the contributions of the individual substituents, but the mesomeric effect of one substituent impacts that of the second and vice versa. For example, the mesomeric interaction is stabilizing if a substituent with a +M effect is in *o*- or *p*- position relative to a substituent that has a −M effect. In this case, the −M substituent withdraws electrons from the same carbon atoms that gain electron density from the +M substituent. Double substituted MAHs with substituents that produce the same type of mesomeric effect (+M or −M) are destabilized if these substituents are in *o*- or *p*- position to each other. Table S9 of the Supporting Information presents all NNIs, which contain significant contributions from mesomeric effects.

Eleven out of 15 NNIs are needed to correct for substituent interactions in ortho position. This indicates the important role of steric or short distance interactions on the thermodynamic properties. Three NNIs describe interactions of substituents in *p*- position and one in *meta* position. In all four cases, mesomeric effects dominate.

**NNI1** accounts for the interaction of a OH group with a CHO group located in *o*- position. The main contribution arises from a very strong hydrogen bond between these groups. The stabilization energy is obtained by second-order perturbation theory analysis to be −23.5 kJ $mol^{-1}$ (see Table S7 of the Supporting Information) and the short distance of 175 pm (1.75 Å) supports this high value. The +M character of CHO and the −M effect of the OH lead to additional stabilization. The presence of these two stabilizing interactions explains the high preliminary value of −28.7 kJ $mol^{-1}$ (see Table S6 of the Supporting Information). The strong hydrogen bond between the OH and the CHO groups restricts the movement of both groups, which is reflected in a reduction of the entropy.

The strong electron withdrawing −M and −I effect of formyl groups are the reason why two formyl groups in ortho position destabilize the molecule. A weak hydrogen bond (calculated stabilization of −6.4 kJ $mol^{-1}$) only partially compensates the destabilizing resonance interaction. Therefore, **NNI2** leads to a large positive enthalpy correction.

**NNI3** describes the repulsion between the lone pairs on the oxygen atoms of two $OCH_3$ groups in *o*- position relative to each other, for example, anomeric effect. The generalized anomeric effect refers to the conformational preference of a gauche structure over an antistructure for molecules with a C—X—C—Y moiety where X and Y are heteroatoms having nonbonding electron pairs (oxygen in this case).[48] Further destabilization results from the combination of +M and −I effects of both groups, hence **NNI3** significantly increases the enthalpy of formation. In triple substituted benzenes, **NNI3** should be used to account for the anomeric effect in the cases where a OH group and a $OCH_3$ are relative to each other in *o*- position and they do not form a hydrogen bond, i.e. lone pairs of the oxygen atoms on OH and $OCH_3$ groups repel each other.

A —CH=$CH_2$ group in ortho position to any other group leads to destabilization. Steric repulsion moves the —CH=$CH_2$ group out of plane and hence reduces the resonance between the aromatic ring and the vinyl group. In **NNI4** and **NNI5**, the neighboring groups are —CHO and —CH=$CH_2$, which both withdraw electrons from the ring via induction and resonance. This leads to additional destabilization.

**NNI6** describes the interaction between an OH group with an $OCH_3$ group in *para* position. Both groups supply via resonance electron density the arene ring, which leads to destabilization as discussed earlier. The electron withdrawing −I effect reduces this destabilization but the mesomeric effects dominates and the overall result is destabilization.

**NNI7** is the interaction of two CHO groups in para position. The combination of strong mesomeric and inductive effects leads to a notable destabilization.

Steric repulsion is the most crucial interaction in **NNI8** which describes the interaction of an alkyl group in ortho position to a formyl group. This leads to a positive $\Delta_f H^\circ$ correction term. As the internal rotation of the alkyl group is hindered, the entropy is reduced.

**NNI9** corrects for the interaction between —CH=$CH_2$ and alkyl groups (—$CH_3$ or —$CH_2CH_3$) in *o*- position. Steric interaction is the only significant factor. The $\Delta_f H^\circ$ correction given for this term is smaller than those for other NNIs (**NNI4** and **NNI5**) that describe ortho vinyl interactions, because no other destabilizing factors are present.

**NNI10** brings a correction term for the case where two CHO substituents are in meta position relative to each other. A CHO substituent withdraws electrons from the phenyl ring via $\sigma$-system regardless of the second substituent group or the position of this second group. As both CHO groups withdraws electrons from the system, electron density is decreased in the ring, which leads to a destabilization. If two CHO groups reside at *o*- or *p*- position relative to each other as in **NNI2** and **NNI7** respectively, −M effect exists and combination of two electron withdrawing effects lead to a stronger destabilization. This is the reason why the enthalpy correction value in **NNI7** is larger than in **NNI10**.

In **NNI11**, a CHO group is in *p*- position to either a OH or a $OCH_3$ group. The combination of +M effect (CHO) with −M effect (OH or $OCH_3$) in *p*- position strongly stabilizes the molecule.

**NNI12** describes the interaction between a CHO and $OCH_3$ group as in o-formylanisole. The combination of the weak hydrogen bond (−3.6 kJ $mol^{-1}$) and the stabilizing mesomeric effect should lead to an overall stabilization. However, the opposite is found, caused by a steric repulsion of both groups. In the most stable conformation, the $OCH_3$ group is directed away from the CHO group. The lone pair electrons of the methoxy group interfere with the $\pi$ electrons of the carbonyl group.

**NNI13** corrects for the steric repulsion between two alkyl groups in *o*- position. Restrictions of the internal rotations cause a reduction of the entropy.

**NNI14** applies to OH groups that are in ortho position to an OH or OR substituent. According to the NBO analysis, the hydrogen bond provides stabilization of −8.9 kJ $mol^{-1}$ (see Table S7 in Supporting Information). The required correction value for the standard enthalpy of formation of **NNI14** is lower, though (see Table S6 of the Supporting Information). This is reasonable because apart from the hydrogen bond, destabilizing +M and −I interactions also play a role. **NNI14** reduces the entropy ($S^\circ$), because the hydrogen bond decreases the flexibility.

**Table 4. Molecules in the Test Set, the Differences Between *Ab Initio* Calculated Data[a] and Data Obtained Using the GAV and NNIs for Standard Enthalpy of Formation ($\Delta_f H°$), Entropies ($S°$) at 298 K, and Enthalpy Capacities ($C_p$) at 300 K Given in Tables S10 and S11 of the Supporting Information and the Statistics for the Test Set**

| Test Set[a] | $\Delta_f H°$ (kJ mol$^{-1}$) | $S°$ (J mol$^{-1}$ K$^{-1}$) | $C_p$ (300 K) (J mol$^{-1}$ K$^{-1}$) |
|---|---|---|---|
| **1,3-divinyl-4-formylbenzene** | −0.1 | 0.9 | 0.4 |
| **1,4-divinyl-2-formylbenzene** | −0.3 | 3.1 | 3.0 |
| **1,3-dimethyl-2-hydroxybenzene** | −1.9 | −3.5 | −1.7 |
| **1,3-dimethyl-5-hydroxybenzene** | −0.4 | 0.4 | 1.4 |
| **1,4-dimethyl-2-hydroxybenzene** | −0.2 | −1.4 | 1.8 |
| **1,2-diethyl-3-hydroxybenzene** | −0.6 | −6.2 | 1.8 |
| **1,3-diethyl-6-hydroxybenzene** | 0.5 | −2.6 | 1.5 |
| **1,4-diethyl-2-hydroxybenzene** | −0.9 | −1.4 | 1.1 |
| **1-ethyl-2-hydroxy-3-formylbenzene** | −2.4 | −1.2 | 3.0 |
| **1,3,5-trimethyl-2-hydroxybenzene** | 0.7 | −4.1 | 0.4 |
| 1,2,3,4-Tetramethylbenzene | 1.0 | 0.7 | −4.9 |
| 1,2,3,4-Tetrahydroxybenzene | −6.5 | 3.6 | −0.3 |
| 1,2-dihydroxy-4,5-diformylbenzene | 0.7 | 1.6 | 5.4 |
| 1-methyl-2-hydroxy-3-methoxy-5-vinylbenzene | −1.2 | −1.3 | −0.7 |
| 1-methyl-3,4,5-trivinylbenzene | −3.2 | −7.3 | −2.1 |
| 1-hydroxy-2-vinyl-4,5-diformylbenzene | −1.9 | −2.1 | −4.8 |
| 1-vinyl-2,4-dimethyl-3-ethylbenzene | 4.0 | 1.4 | −6.8 |
| 1,3-dimethoxy-2-hydroxy-5-methylbenzene | 1.8 | 3.1 | −1.4 |
| 1,3,4-trihydroxy-2-methoxybenzene | 2.0 | −1.7 | −1.9 |
| 1,3-vinyl-2-hydroxy-4-methylbenzene | −1.8 | −5.0 | −3.4 |
| 1-hydroxy-2-formyl-4-methyl-5-vinylbenzene | 1.5 | 0.7 | 0.5 |
| 1,2-dihydroxy-3-methoxy-5-ethylbenzene | −1.1 | 7.2 | −4.0 |
| 1,2-dihydroxy-3-methoxy-5,6-dimethylbenzene | −0.5 | −3.2 | −5.7 |
| 1-hydroxy-2,3-diformyl-4-methyl-6-vinylbenzene | −2.0 | −7.1 | −2.4 |
| 1,2,3,4,5-Pentahydroxybenzene | −8.7 | 4.4 | 0.1 |
| **Statistics for the Test Set** | | | |
| SD | 2.5 | 3.6 | 2.9 |
| MAD | 1.8 | 3.0 | 2.4 |
| MAX | 8.7 | 7.3 | 6.8 |

[a]Ten Molecules for which data is collected at G4 are given in boldface.

If styrene is substituted in *o*- position by —OH or —$OCH_3$, which are +M groups, several effects occur simultaneously. Steric repulsion increases the enthalpy but the +M/−I combination provides stabilization. In total, a small destabilization results.

### *Simultaneous determination and validation of GAVs and NNIs*

In earlier sections, preliminary values were assigned to the GAVs and these GAVs were used to reproduce the thermochemical data of double substituted benzenes. The deviations between the calculations with these preliminary GAVs and the reference data were used to identify 15 NNIs. Then, physical interpretations of these interactions were provided with the help of NBO calculations. In this section, these GAV and NNI parameters are optimized simultaneously via least squares fitting using the information of the entire training set. These GAVs and NNIs are given in Supporting Information Tables S10 and S11, respectively. Performance of the GAV/NNI parameter set is assessed through a statistical analysis. The results of this analysis are reported in Supporting Information Table S12. The MAD for the GA calculation of the standard enthalpies of formation of the entire reference data set of 133 molecules is 0.95 kJ mol$^{-1}$, and those for group additively calculated entropy and heat capacity data at all temperatures are 1.60 J mol$^{-1}$ K$^{-1}$ and $<$ 1.56 J mol$^{-1}$ K$^{-1}$, respectively. The improvement on the inclusion of non-nearest neighbor NNIs is significant for all three thermodynamic properties. Introduction of NNIs lower the MAD for $\Delta_f H°$ from 5.70 to 0.95 kJ mol$^{-1}$, and similar improvement can also be noted in the group additively calculated data for entropies (from 3.9 to 1.6 J mol$^{-1}$ K$^{-1}$) and heat capacities (from 4.0 to 1.6 J mol$^{-1}$ K$^{-1}$).

To test the transferability of the GAVs and NNIs, thermodynamic properties of a test set consisting of 9 triple substituted, 13 quadruple-substituted and 3 quintuple substituted MAHs are obtained using the GAVs and NNIs reported in Tables S10 and S11 of the Supporting Information. Due to the higher computational cost for the G4 method, *ab initio* data for 12 quadruple-substituted and 3 quintuple substituted MAHs could only be obtained with the CBS-QB3/BAC method whereas for the other 10 molecules, *ab initio* data is obtained with the G4/BAC method. As it has been shown that CBS-QB3 and G4 perform with similar accuracy for closed shell MAHs, reproducing CBS-QB3 *ab initio* data with GAVs obtained from G4 database is considered as an appropriate practice.

Differences between the *ab initio* and the GA based data are provided in Table 4 and the actual *ab initio* values for $\Delta_f H°$, S°, and $C_p$ are collected in Table S13 of the Supporting Information. The statistics of the test set in Table 4 demonstrates the good performance of the GAV/NNI set developed in this study. The MAD values of 1.8 kJ mol$^{-1}$ ($\Delta_f H°$) and 3.0 J mol$^{-1}$ K$^{-1}$ and 2.4 J mol$^{-1}$ K$^{-1}$ for $S°$ and $C_p$(300 K), respectively, prove that the group values are transferable and that even highly substituted benzenes with strongly interacting substituents are obtained by GA with high accuracy. Since only 2 of the 15 NNIs have been reported in the literature before, data obtained with the GAVs and available NNIs in the literature is insufficient in reproducing thermochemical data for the multiple substituted benzenes. In this manner, the reported set of GAVs/NNIs enables the application of the method of Benson's GA to obtain thermochemical data for multiple substituted benzenes even for the quadruple and

**Table 5. GAVs for Standard Enthalpy of Formation ($\Delta_f H^\circ$) and Entropy ($S^\circ$) at 298 K, and Heat Capacity ($C_p$) at Various Temperatures for MAHs Derived from Reference Database, Given with 97.5% Confidence Intervals**

| | | | $C_p$ (J mol$^{-1}$ K$^{-1}$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| GAVs | $\Delta_f H^\circ$ (kJ mol$^{-1}$) | $S^\circ$ (J mol$^{-1}$ K$^{-1}$) | 300 K | 400 K | 500 K | 600 K | 800 K | 1000 K | 1500 K |
| $C_b$—(H) | 13.7 ± 0.16 | 48.6 ± 0.28 | 13.3 ± 0.26 | 18.4 | 22.7 | 26.2 | 31.3 | 34.8 | 40.1 |
| $C_b$—(O) | 24.0 ± 0.41 | −30.0 ± 0.70 | 12.1 ± 0.66 | 12.6 | 15.0 | 17.4 | 21.3 | 23.5 | 26.0 |
| O—($C_b$)C | −124.1 ± 0.51 | 21.6 ± 0.88 | 20.3 ± 0.82 | 24.2 | 25.1 | 24.8 | 23.7 | 23.1 | 22.0 |
| $C_b$—(CO) | 21.5 ± 0.73 | −33.5 ± 1.25 | 16.4 ± 1.16 | 17.8 | 19.7 | 21.6 | 24.5 | 26.6 | 27.4 |
| $C_b$—($C_d$) | 24.3 ± 0.54 | −35.0 ± 0.92 | 12.4 ± 0.86 | 14.8 | 16.4 | 18.0 | 20.9 | 22.6 | 25.6 |
| $C_b$—(C) | 23.5 ± 0.40 | −33.6 ± 0.68 | 10.8 ± 0.63 | 13.8 | 16.5 | 18.7 | 21.9 | 23.8 | 26.2 |
| C—($C_b$)(C)($H$)$_2$ | −20.6 ± 0.45 | 38.6 ± 0.76 | 24.9 ± 0.71 | 30.7 | 36.0 | 40.4 | 47.3 | 52.3 | 60.0 |

quintuple substituted benzenes that possess substituent groups of interest.

As a final step, the aforementioned 10 molecules in the test set for which *ab initio* data are collected at the G4 level is added into the training set of 133 molecules. Then, GAV and NNI parameters are optimized through a final linear regression analysis based on this reference dataset of 143 molecules. The final GAV and NNI parameters together with their 97.5% confidence intervals are reported in Table 5 (GAVs) and Table 6 (NNIs). Only the confidence intervals for heat capacities ($C_p$) at 300 K are provided—those for the other $C_p$ data have similar or lower values.

GAVs for singly substituted MAHs have been derived before by Benson et al.,[15] Cohen,[17] Holmes and Aubry,[19] and Sabbe et al.[22] The present GAVs for enthalpy of formation are compared in Table 7 to those previously derived values. At a first glance, it seems that there are large deviations between the current data and those reported by Benson, Cohen, and Holmes, for example, for the O—($C_b$)(C) group. However, this can be explained by the fact that Benson and Cohen use a different way to eliminate linear dependencies than Holmes, Sabbe, and this study. This illustrates the importance to use only internally consistent GAVs to obtain thermodynamic properties. This also shows that direct comparison of GAVs is not feasible. To allow a fair comparison, Table 7 also list the sum of GAV values for pairs of linear dependent groups. From Table 7, it can be seen that the sum of linear dependent GAVs $C_b$—(O) and O—($C_b$)(H) obtained by Cohen, Holmes, and the GAVs reported in this study agree well. In earlier sections, it has been mentioned that the experimental heat of formation for anisole (methoxybenzene) is not well established. This uncertainty is reflected in the sum of the linear GAVs $C_b$—(O) and O—($C_b$)(C). The value reported by Cohen for this pair is significantly higher than the other data, because his GAV was derived from the highest experimental enthalpy of formation for anisole (Fenwick et al.).[44] The assignments by Benson and Cohen and this study for GAV of the linear dependent pair $C_b$—(CO)+CO—($C_b$)(H) match very well, whereas the value reported by Holmes is slightly lower. Good agreement is also noted for the pairs of linear dependent GAVs that include the $C_b$—(C) group. Overall, the final GAV assignments derived in this work are very consistent with previous studies despite the fact that a significantly larger data set was used and that the GAVs were derived via full optimization together with the NNIs. This means that the selection of NNIs did not cause any notable interference with the GAVs.

Holmes and Aubry[19] investigated cis and trans 1-hydroxypropenal at the CBS-QB3 level of theory and explained the observed stability difference with intramolecular interactions that are very similar to **NNI1**. The authors report a CBS-QB3 $\Delta_f H^\circ$ value of −241 kJ mol$^{-1}$ for the trans isomer and a value of −276 kJ mol$^{-1}$ for the cis isomer of 1-hydroxypropenal. They mention the fact that GA data for $\Delta_f H^\circ$ of the trans isomer agrees well (−246 ± 4 kJ mol$^{-1}$) with the reported CBS-QB3 data, but to obtain the enthalpy of formation of the more stable cis isomer accurately, an additional correction of −30 kJ mol$^{-1}$ is needed. This value is in good agreement with the **NNI1** derived in this work. To correct for the interaction between two —$OCH_3$ groups, Cohen has suggested a $\Delta_f H^\circ$ correction of 12.6 kJ mol$^{-1}$ which is

**Table 6. Corrections for NNIs Derived Based on Full G4 Molecule Set for the Standard Enthalpies of Formation ($\Delta_f H^\circ$) and Entropies ($S^\circ$) at 298 K and Heat Capacities ($C_p$) at Various Temperatures for MAHs[a]**

| | | | | $C_p$ (J mol$^{-1}$ K$^{-1}$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| NNI | Interaction[b] | $\Delta_f H^\circ$(kJ mol$^{-1}$) | $S^\circ$ (J mol$^{-1}$ K$^{-1}$) | 300 K | 400 K | 500 K | 600 K | 800 K | 1000 K | 1500 K |
| **NNI1** | o-OH+CHO | −27.4 ± 1.52 | −21.3 ± 2.60 | −10.4 ± 2.43 | −9.0 | −7.4 | −6.0 | −3.4 | −1.4 | 1.8 |
| **NNI2** | o-CHO+CHO | 21.1 ± 1.37 | 6.4 ± 2.35 | 1.7 ± 2.19 | 1.7 | 1.6 | 1.4 | 0.4 | −0.8 | −2.5 |
| **NNI3** | Anomeric Effect | 14.7 ± 1.25 | 7.8 ± 2.14 | 0.0 ± 2.00 | −3.4 | −4.7 | −5.0 | −4.4 | −3.7 | −2.2 |
| **NNI4** | o-CH=$CH_2$+CHO | 11.9 ± 1.77 | −2.6 ± 3.03 | 1.3 ± 2.82 | 1.6 | 2.1 | 2.5 | 2.4 | 1.8 | 0.3 |
| **NNI5** | o-CH=$CH_2$+CH=$CH_2$ | 8.1 ± 1.13 | −2.3 ± 1.93 | 4.6 ± 1.80 | 3.2 | 2.1 | 1.3 | 0.2 | −0.2 | −0.5 |
| **NNI6** | p-OH/MeO+OH/MeO | 7.3 ± 1.24 | 4.0 ± 2.12 | 1.7 ± 1.98 | 0.4 | −0.8 | −1.7 | −2.3 | −2.3 | −1.7 |
| **NNI7** | p-CHO+CHO | 9.9 ± 1.99 | −0.8 ± 3.40 | 1.8 ± 3.17 | 1.4 | 1.0 | 0.7 | 0.1 | −0.3 | −0.9 |
| **NNI8** | o-Me/Et+CHO | 8.1 ± 1.48 | −2.4 ± 2.53 | 4.0 ±2.36 | 3.0 | 2.0 | 1.1 | 0.0 | −0.6 | −0.9 |
| **NNI9** | o-CH=$CH_2$+Me/Et | 4.6 ± 1.29 | −5.7 ± 2.21 | 5.7 ± 2.06 | 5.6 | 4.9 | 4.2 | 2.9 | 2.1 | 1.0 |
| **NNI10** | m-CHO+CHO | 4.8 ± 1.13 | 0.1 ± 1.94 | 1.7 ± 1.80 | 1.0 | 0.5 | 0.1 | −0.4 | −0.7 | −0.8 |
| **NNI11** | p-CHO+OH/MeO | −4.6 ± 1.30 | −0.8 ± 2.23 | −1.6 ± 2.08 | −1.0 | −0.3 | 0.2 | 0.9 | 1.2 | 1.5 |
| **NNI12** | o-MeO+CHO | 7.9 ± 1.73 | −1.7 ± 2.95 | −2.6 ± 2.75 | 0.3 | 2.6 | 3.9 | 4.1 | 3.3 | 1.3 |
| **NNI13** | o-Me/Et+Me/Et | 4.2 ± 0.86 | −6.6 ± 1.47 | 3.3 ± 1.37 | 2.9 | 2.6 | 2.5 | 2.3 | 2.0 | 1.5 |
| **NNI14** | o-OH+OH/MeO | −3.0 ± 0.78 | −5.6 ± 1.34 | −2.7 ± 1.25 | −1.5 | 0.2 | 2.0 | 4.2 | 4.8 | 3.7 |
| **NNI15** | o-CH=$CH_2$+OH/MeO | 2.6 ± 1.20 | −3.3 ± 2.04 | 2.5 ± 1.91 | 2.7 | 2.6 | 2.5 | 2.1 | 1.8 | 1.2 |

[a]97.5% confidence intervals for $\Delta_f H^\circ$ (298 K), $S^\circ$ (298 K), and $C_p$ (300 K) are provided.
[b]Substituent groups are given as OH: hydroxy (—OH), MeO: methoxy (—$OCH_3$), Formyl: —CHO, Vinyl: —CH=$CH_2$, Me: methyl (—$CH_3$), Et: ethyl (—$CH_2CH_3$).

**Table 7. GAVs and Pairs of Linearly Dependent GAVs for the Standard Enthalpies of Formation ($\Delta_f H^\circ$) Reported in the Literature and Obtained in this Work**

| | Literature | | | | |
|---|---|---|---|---|---|
| GAVs | Benson et al.[15] | Cohen[18] | Holmes and Aubry[19] | Sabbe et al./ Paraskevas[22,25] | This Work |
| **$\Delta_f H^\circ$ (kJ mol$^{-1}$)** | | | | | |
| $C_b$—(H) | 13.8 | 13.8 | 13.8 | 13.8 | 13.7 |
| $C_b$—(O) | −3.8 | −3.8 | 25.0 | – | 24.0 |
| O—($C_b$)(C) | −96.2 | −90.4 | −124.0 | – | −124.1 |
| $C_b$—(CO) | 15.5 | 15.5 | 18.0 | – | 21.5 |
| $C_b$—($C_d$) | 23.8 | 24.3 | 27.0 | 24.0 | 24.3 |
| $C_b$—(C) | 23.1 | 23.0 | 23.0 | 24.4 | 23.5 |
| C—($C_b$)(C)$(H)_2$ | −20.3 | −19.2 | −20.0 | −21.2 | −20.6 |
| O—($C_b$)(H) | −158.6 | −161.1 | −190.0 | – | −188.1[a] |
| C—(O)$(H)_3$ | −42.2 | −42.9 | −42.0 | −42.9 | −42.9[b] |
| CO—($C_b$)(H) | −121.8 | −122 | −122.0 | – | −128.3[a] |
| $C_d$—($C_b$)(H) | 28.4 | 26.5 | 28.0 | 30.4 | 30.4[b] |
| $C_d$—$(H)_2$ | 26.2 | 26.4 | 26.2 | 25.1 | 25.1[b] |
| C—($C_b$)$(H)_3$ | −42.7 | −42.0 | −42.0 | −42.9 | −42.9[b] |
| C—(C)$(H)_3$ | −42.7 | −42.0 | −42.0 | −42.9 | −42.9[b] |
| **Pairs of Linear Dependent GAVs** | | | | | |
| $C_b$—(O) + O—($C_b$)(H) | −162.4 | −164.9 | −165.0 | – | −164.1 |
| $C_b$—(O) + O—($C_b$)(C) | −100.0 | −94.2 | −99.0 | – | −100.1 |
| $C_b$—(CO) + CO—($C_b$)(H) | −106.3 | −106.5 | −104 | – | −106.8 |
| $C_b$—($C_d$) + $C_d$—($C_b$)(H) | 52.1 | 50.8 | 55.0 | 54.4 | 54.7 |
| $C_b$—(C) + C—($C_b$)$(H)_3$ | −19.6 | −19.0 | −19.0 | −19.9 | −19.4 |
| $C_b$—(C) + C—($C_b$)(C)$(H)_2$ | 2.8 | 3.8 | 3 | 1.8 | 2.9 |

[a]Values of these GAVs are set equal to the values of structurally GAVs reported in the previous studies of Sabbe and coworkers[22,25] to eliminate linear dependencies.
[b]Values of these GAVs are taken from the previous studies of Sabbe and coworkers[22,25] to eliminate linear dependencies.

**Table 8. Statistics for the Linear Regression Analysis of the GAVs and NNIs for the Standard Enthalpies of Formation ($\Delta_f H^\circ$) and Entropies ($S^\circ$) at 298 K and Heat Capacities ($C_p$) at Various Temperatures for MAHs[a]**

| | | | $C_p$ (J mol$^{-1}$ K$^{-1}$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\Delta_f H^\circ$ (kJ mol$^{-1}$) | $S^\circ$ (J mol$^{-1}$ K$^{-1}$) | 300 K | 400 K | 500 K | 600 K | 800 K | 1000 K | 1500 K |
| F | 33,916 | 19,264 | 13,165 | 29,262 | 53,571 | 79,914 | 122,625 | 160,869 | 345,956 |
| MAD | 0.93 | 1.66 | 1.56 | 1.32 | 1.15 | 1.06 | 0.97 | 0.93 | 0.72 |
| RMS | 1.24 | 2.12 | 1.97 | 1.68 | 1.48 | 1.37 | 1.30 | 1.25 | 0.96 |
| MAX | 4.42 | 7.03 | 5.65 | 4.54 | 4.65 | 4.48 | 4.00 | 4.05 | 3.73 |

[a]F, Significance; MAD, mean average deviation; RMS, root mean square; MAX, maximum deviation.

comparable to **NNI3** (14.7 kJ mol$^{-1}$) reported in the present study. The derived $\Delta_f H^\circ$ correction for **NNI6** is found to be 4.1 kJ mol$^{-1}$ for the interaction of two alkyl groups in *o*- position to each other which is expected to be a similar interaction to the steric repulsion between two alkyl groups in cis alkenes. In the work of Sabbe et al.,[22] a NNI of 5.9 kJ mol$^{-1}$ is assigned for the latter interaction, while a value of 4.6 kJ mol$^{-1}$ was obtained by Cohen[18] and Benson[16] reports 4.2 kJ mol$^{-1}$.In this study, the **NNI6** correction term for $S^\circ$ is derived as −6.2 J mol$^{-1}$ K$^{-1}$ which is close to the "ortho" corrections derived earlier: −5.7 J mol$^{-1}$ K$^{-1}$ by Sabbe et al.,[24] −6.7 J mol$^{-1}$ K$^{-1}$ by Benson et al.[15] and Cohen.[18] All these comparisons demonstrate that the derived NNIs are plausible.

A statistical analysis is performed to assess the performance of the final GA parameters. The results are reported in Table 8. The MAD between the GA and *ab initio* calculated data of the standard enthalpies of formation of the entire reference data set of 143 molecules is 0.93 kJ mol$^{-1}$, and those for the entropy and heat capacity data at all temperatures are 1.66 J mol$^{-1}$ K$^{-1}$ and $< 1.56$ J mol$^{-1}$ K$^{-1}$, respectively. Next-nearest neighbor interactions (NNIs) are found to be crucial as the drastic decrease of the MAD values shows. For example,

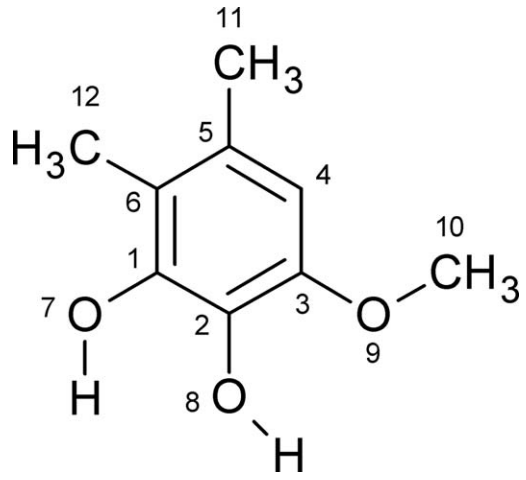


**Figure 1. Structure of 1,2-dihydroxy-3-methoxy-5,6-dimethylbenzene given with numbers of the atom centers of the groups.**

**Table 9. Application of GAVs/NNIs for the Calculation of Standard Enthalpy of Formation ($\Delta_f H^\circ$) and Entropy ($S^\circ$) of 1,2-Dihydroxy-3-Methoxy-5,6-Dimethylbenzene (see Figure 1)**

| Application of Group Additive Method | | | | | | | |
|---|---|---|---|---|---|---|---|
| GAVs | | | | NNIs | | | |
| Center # | GAVs | $\Delta_f H^\circ$ (kJ mol$^{-1}$) | $S^\circ$ (J mol$^{-1}$ K$^{-1}$) | NNI | Interaction | $\Delta_f H^\circ$ (kJ mol$^{-1}$) | $S^\circ$ (J mol$^{-1}$ K$^{-1}$) |
| 1 | $C_b$—(O) | 24.0 | −30.0 | Interactions of Substituent OH (Atom 7) | | | |
| 2 | $C_b$—(O) | 24.0 | −30.0 | **NNI14** | *o*-OH+OH | −3.0 | −5.6 |
| 3 | $C_b$—(O) | 24.0 | −30.0 | – | *m*-OH+ $OCH_3$ | 0 | 0 |
| 4 | $C_b$—(H) | 13.7 | 48.6 | – | *o*-OH+Me | 0 | 0 |
| 5 | $C_b$—(C) | 23.5 | −33.6 | – | *m*-OH+Me | 0 | 0 |
| 6 | $C_b$—(C) | 23.5 | −33.6 | Interactions of Substituent OH (Atom 8) | | | |
| 7 | O—($C_b$)(H) | −188.1 | 106.3 | **NNI14** | *o*-OH+ $OCH_3$ | −3.0 | −5.6 |
| 8 | O—($C_b$)(H) | −188.1 | 106.3 | Interactions of Substituent $OCH_3$ (Atoms 9 and 10) | | | |
| 9 | O—($C_b$)(C) | −124.1 | 21.6 | – | *m*- $OCH_3$+Me | 0 | 0 |
| 10 | C—(O)($H_3$) | −42.9 | 127.1 | – | *p*- $OCH_3$+Me | 0 | 0 |
| 11 | C—($C_b$)($H_3$) | −42.9 | 127.2 | Interactions of Substituent Me (Atom 11) | | | |
| 12 | C—($C_b$)($H_3$) | −42.9 | 127.2 | **NNI13** | *o*-Me+Me | +4.2 | −6.6 |
| $\sum$ GAVs | | −496.3 | 507.2 | $\sum$ NNIs | | −1.8 | −17.8 |
| $\sum$ GAVs+NNIs | | −506.5 | 489.4 | | | | |
| Standard Intrinsic Entropy Calculation | | | | Comparison of CBS-QB3 vs. GAVs/NNIs | | | |
| $\sigma_{int}$ | 3*3*3=27 (Atoms 10, 11, 12) | $\sigma/n_{opt}$ | 27 | | | $\Delta_f H^\circ$ (kJ mol$^{-1}$) | $S^\circ$ (J mol$^{-1}$ K$^{-1}$) |
| $\sigma_{ext}$ | 1 | $S^\circ$(CBS-QB3) | 478.9 J mol$^{-1}$ K$^{-1}$ | CBS-QB3 | | −498.5 | 486.7 |
| $\sigma$ | 27 | $S^o_{int}$ (CBS-QB3) | 486.7 J mol$^{-1}$ K$^{-1}$ | $\sum$ GAVs+NNIs | | −498.1 | 489.4 |
| $n_{opt}$ | 1 | | | (CBS-QB3) − $\sum$ GAVs+NNIs | | −0.4 | −2.7 |

the MAD for $\Delta_f H^\circ$ decreases from 5.70 to 0.95 kJ mol$^{-1}$ after NNI corrections are introduced.

All individual deviations between *ab initio* and group additively calculated thermodynamic data are provided in Table S14 of the Supporting Information. The distribution of the difference between the reference and the GA calculated enthalpies of formation is shown in **Figure S4** of Supporting Information. Almost all of the 143 enthalpies of formation obtained by GA are within 4 kJ mol$^{-1}$. Marginally higher deviations are seen for just two molecules: 4.1 kJ mol$^{-1}$ for 1-hydroxy-2-methoxy-4-formylbenzene (**molecule 99**) and 4.4 kJ mol$^{-1}$ for 1-formyl-3-methoxy-4-ethylbenzene (**molecule 124**). In **Figures S5** and **S6** of Supporting Information, similar histograms are presented for the difference distributions of the entropies ($S^\circ$) and heat capacities ($C_p$) at 300 K. Again, the agreement is very good and most data differ by less than 4 J mol$^{-1}$ K$^{-1}$.

### *Application of GAVs and NNIs*

To demonstrate the use of the GAVs (see Table 4) and the NNIs (see Table 5) reported are in this study, an example calculation is done to obtain $\Delta_f H^\circ$ and $S^\circ$ of 1,2-dihydroxy-3-methoxy-5,6-dimethylbenzene. The structure of this molecule is given in Figure 1.

The 12 GAVs and 3 NNIs required for the GA calculation of standard enthalpy of formation ($\Delta_f H^\circ$) and entropy ($S^\circ$) of this quintuple substituted MAH are shown in Table 9. **NNI 13** is used to correct for the interaction between two methyl groups in *o*- position to each other; **NNI14** is used twice to take two other *o*- interactions into account: the interaction between two OH groups and the interaction a OH and a $OCH_3$ group. No NNIs have been defined for other less significant interactions in this molecule which are also indicated in Table 9. The difference between CBS-QB3 values and GA calculated data for $\Delta_f H^\circ$ and $S^\circ$ are −0.4 and −2.7 J mol$^{-1}$ K$^{-1}$ which shows that the neglected interactions do not lead to a significant loss of accuracy in the group additively calculated data and the present GAVs/NNIs are successful in describing thermochemical data for this quintuple substituted benzene.

## Conclusions

A set of seven GAV and 15 NNI has been derived to obtain the thermodynamic properties of MAHs with six substituents: hydroxy (—OH), methoxy (—$OCH_3$), formyl (—CHO), vinyl (—CH=$CH_2$) and the alkyl groups methyl (—$CH_3$) and ethyl (—$CH_2CH_3$). The GAV and NNI parameters were determined from a set of 143 molecules (reference data), whose thermodynamic values were calculated with bond additivity corrected (BAC) G4 theory. The BAC values were obtained as part of this study from a set of 77 molecules, for which accurate experimental data exist. Parallel computations with the CBS-QB3 method yielded similar results, which indicates that the CBS-QB3 method is reliable for closed-shell molecules and observed problems with bond dissociation energies are caused by problems to describe the phenyl radical accurately.

NNI were found to be crucial to achieve accurate GA calculations. NBO calculations were employed to ensure that the defined NNIs are meaningful and to provide insight into the type of interactions that cause either stabilization or destabilization. Values obtained with the optimized GAV and NNI parameters agree very well with the reference data set. The MAD are 0.93 kJ mol$^{-1}$ for $\Delta_f H^\circ$, 1.66 J mol$^{-1}$ K$^{-1}$ for $S^\circ$, and less than 1.56 J mol$^{-1}$ K$^{-1}$ for $C_p$ values at all studied temperatures. All GA based enthalpies are within 5 kJ mol$^{-1}$ for the *ab initio* values and only for two molecules, the deviation exceeds 4 kJ mol$^{-1}$. Similarly, the entropies are for almost all molecules reproduced to within 4 J mol$^{-1}$ K$^{-1}$, which demonstrated the high accuracy obtained with the GAV/NNI set calculated in this study. Further validation using

nine triple-substituted, thirteen quadruple-substituted, and one quintuple-substituted benzene underscores the predictive capability. Combined with previously published GAV for hydrocarbon and oxygenated hydrocarbon species, an internally consistent set of GA parameters is now available for closed-shell monocyclic aromatic molecules. Future work will extend this toward radicals.

## Acknowledgments

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013)/ERC grant agreement n° 290793. This work was carried out using the STEVIN Supercomputer Infrastructure at Ghent University, funded by Ghent University, the Flemish Supercomputer Center (VSC), and the Hercules Foundation and the Flemish Government—department EWI.

## Literature Cited

1. Clymans PJ, Froment GF. Computer-generation of reaction paths and rate-equations in the thermal-cracking of normal and branched paraffins. *Comput Chem Eng*. 1984;8(2):137–142.
2. Hillewaert LP, Dierickx JL, Froment GF. Computer-generation of reaction schemes and rate-equations for thermal-cracking. *AIChE J*. 1988;34(1):17–24.
3. Green WH, Barton PI, Bhattacharjee B, Matheu DM, Schwer DA, Song J, Sumathi R, Carstensen H-H, Dean AM, Grenda JM. Computer construction of detailed chemical kinetic models for gas-phase reactors. *Ind Eng Chem Res*. 2001;40(23):5362–5370.
4. Sabbe MK, Van Geem KM, Reyniers MF, Marin GB. First principle-based simulation of ethane steam cracking. *AIChE J*. 2011; 57(2):482–496.
5. Ranzi E. A wide-range kinetic modeling study of oxidation and combustion of transportation fuels and surrogate mixtures. *Energy Fuels*. 2006;20(3):1024–1032.
6. Tran LS, Sirjean B, Glaude P-A, Fournet R, Battin-Leclerc F. Progress in detailed kinetic modeling of the combustion of oxygenated components of biofuels. *Energy*. 2012;43(1):4–18.
7. Richter H, Risoul V, Lafleur AL, Plummer EF, Howard JB, Peters WA. Chemical characterization and bioactivity of polycyclic aromatic hydrocarbons from non-oxidative thermal treatment of pyrene-contaminated soil at 250-1,000 degrees C. *Environ Health Perspect*. 2000;108(8):709–717.
8. Norinaga K, Deutschmann O, Saegusa N, Hayashi J-I. Analysis of pyrolysis products from light hydrocarbons and kinetic modeling for growth of polycyclic aromatic hydrocarbons with detailed chemistry. *J Anal Appl Pyrolysis*. 2009;86(1):148–160.
9. Saggese C, Sánchez NE, Frassoldati A, Cuoci A, Faravelli T, Alzueta MU, Ranzi E. Kinetic modeling study of polycyclic aromatic hydrocarbons and soot formation in acetylene pyrolysis. *Energy Fuels*. 2014;28(2):1489–1501.
10. Evans RJ, Milne TA. Molecular characterization of the pyrolysis of biomass. *Energy Fuels*. 1987;1(2):15.
11. Patwardhan PR, Brown RC, Shanks BH. Understanding the fast pyrolysis of lignin. *ChemSusChem*. 2011;4(11):1629–1636.
12. Sangha AK, Petridis L, Smith JC, Ziebell A, Parks JM. Molecular simulation as a tool for studying lignin. *Environ Prog Sustain Energy*. 2012;31(1):47–54.
13. Benson SW, Buss JH. Additivity rules for the estimation of molecular properties—thermodynamic properties. *J Chem Phys*. 1958;29(3): 546–572.
14. Benson SW. *Thermochemical Kinetics; Methods for the Estimation of Thermochemical Data and Rate Parameters*. New York: Wiley; 1968.
15. Benson SW, Cruickshank FR, Golden DM, Haugen G, O'Neal HE, Rodgers AS, Shaw R, Walsh R. Additivity rules for the estimation of thermochemical properties. *Chemical Rev*. 1969;69(3): 279–324.
16. Benson SW. *Thermochemical Kinetics: Methods for the Estimation of Thermochemical Data and Rate Parameters*, *2nd ed*. New York: Wiley; 1976.
17. Cohen N. Thermochemistry of alkyl free-radicals. *J Phys Chem*. 1992;96(22):9052–9058.
18. Cohen N. Revised group additivity values for enthalpies of formation (at 298 K) of carbon–hydrogen and carbon–hydrogen–oxygen compounds. *J Phys Chem Ref Data*. 1996;25(6):1411.
19. Holmes JL, Aubry C. Group additivity values for estimating the enthalpy of formation of organic compounds: an update and reappraisal. 1. C, H, and O. *J Phys Chem A*. 2011;115(38):10576–10586.
20. Holmes JL, Aubry C. Group Additivity values for estimating the enthalpy of formation of organic compounds: an update and reappraisal. 2. C, H, N, O, S, and Halogens. *J Phys Chem A*. 2012; 116(26):7196–7209.
21. Marsi I, Viskolcz B, Seres L. Application of the group additivity method to alkyl radicals: an ab initio study. *J Phys Chem A*. 2000; 104(19):4497–4504.
22. Sabbe MK, Saeys M, Reyniers M-F, Marin GB, Van Speybroeck V, Waroquier M. Group additive values for the gas phase standard enthalpy of formation of hydrocarbons and hydrocarbon radicals. *J Phys Chem A*. 2005;109(33):7466–7480.
23. Sabbe MK, De Vleeschouwer F, Reyniers M-F, Waroquier M, Marin GB. First principles based group additive values for the gas phase standard entropy and heat capacity of hydrocarbons and hydrocarbon radicals. *J Phys Chem A*. 2008;112(47):12235–12251.
24. Vandeputte AG, Sabbe MK, Reyniers MF, Marin GB. Modeling the gas-phase thermochemistry of organosulfur compounds. *Chemistry*. 2011;17(27):7656–7673.
25. Paraskevas PD, Sabbe MK, Reyniers M-F, Papayannakos N, Marin GB. Group additive values for the gas-phase standard enthalpy of formation, entropy and heat capacity of oxygenates. *Chemistry*. 2013;19(48):16431–16452.
26. Sumathi R, Green W. Missing thermochemical groups for large unsaturated hydrocarbons: contrasting predictions of G2 and CBS-Q. *J Phys Chem A*. 2002;106(46):11141–11149.
27. Sun H, Bozzelli JW. Structures, intramolecular rotation barriers, and thermochemical properties: ethanol, α-monoethanols, dichloroethanols, and corresponding radicals derived from H atom loss. *J Phys Chem A*. 2001;105(41):9543–9552.
28. Sun H, Bozzelli JW. Structures, rotational barriers, thermochemical properties, and additivity groups for 2-propanol, 2-chloro-2-propanol and the corresponding alkoxy and hydroxyalkyl radicals. *J Phys Chem A*. 2002;106(15):3947–3956.
29. *Gaussian 09*, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr., Peralta JE, Ogliaro F, Bearpark MJ, Heyd J, Brothers EN, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam NJ, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ, Gaussian, Inc.: Wallingford, CT, USA, 2009.
30. Montgomery JA Jr., Frisch MJ, Ochterski JW, Petersson GA. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J Chem Phys*. 1999;110(6):2822–2827.
31. Curtiss LA, Redfern PC, Raghavachari K. Gaussian-4 theory. *J Chem Phys*. 2007;126(8):084108.
32. Paraskevas PD, Sabbe MK, Reyniers MF, Papayannakos N, Marin GB. Kinetic modeling of alpha-hydrogen abstractions from unsaturated and saturated oxygenate compounds by carbon-centered radicals. *ChemPhysChem*. 2014;15(9):1849–1866.
33. Paraskevas PD, Sabbe MK, Reyniers M-F, Papayannakos NG, Marin GB. Kinetic modeling of α-hydrogen abstractions from unsaturated and saturated oxygenate compounds by hydrogen atoms. *J Phys Chem A*. 2014;118(40):9296–9309.
34. Sabbe M, Vandeputte A, Reyniers M-F, Van Speybroeck V, Waroquier M, Marin G. Ab initio thermochemistry and kinetics for carbon-centered radical addition and β-scission reactions. *J Phys Chem A*. 2007;111(34):8416–8428.
35. Sabbe M, Reyniers M-F, Van Speybroeck V, Waroquier M, Marin G. Carbon-centered radical addition and β-scission reactions: modeling of activation energies and pre-exponential factors. *ChemPhysChem*. 2008;9(1):124–140.

36. Luo Y. Bond dissociation energies. In Lide DR, editor. *CRC Handbook of Chemistry and Physics*. Boca Raton, FL: CRC Press/Taylor and Francis; 2008.
37. Blanksby SJ, Ellison GB. Bond dissociation energies of organic molecules. *Acc Chem Res*. 2003;36(4):255–263.
38. Vansteenkiste P, Van Speybroeck V, Marin GB, Waroquier M. Ab initio calculation of entropy and heat capacity of gas-phase n-alkanes using internal rotations. *J Phys Chem A*. 2003;107(17):3139–3145.
39. Van Speybroeck V, Vansteenkiste P, Neck DV, Waroquier M. Why does the uncoupled hindered rotor model work well for the thermodynamics of n-alkanes? *Chem Phys Lett*. 2005;402(4–6):479–484.
40. Vansteenkiste P, Van Neck D, Van Speybroeck V, Waroquier M. An extended hindered-rotor model with incorporation of Coriolis and vibrational-rotational coupling for calculating partition functions and derived quantities. *J Chem Phys*. 2006;124(4):044314.
41. Curtiss LA, Raghavachari K, Redfern PC, Pople JA. Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *J Chem Phys*. 1997;106(3):1063–1079.
42. Petersson GA, Malick D, Wilson W, Ochterski J, Montgomery JA, Frisch MJ. Calibration and comparison of the Gaussian-2, complete basis set, and density functional methods for computational thermochemistry. *J Chem Phys*. 1998;109(24):10570.
43. Burgess DR. *"Thermochemical Data" in NIST Chemistry WebBook, NIST Standard Reference Database Number 69*. Gaithersburg, MD: National Institute of Standards and Technology; 2005.
44. Fenwick J, Harrop D, Head A. Thermodynamic properties of organic oxygen compounds 41. Enthalpies of formation of eight ethers. *J Chem Thermodyn*. 1975;7(10):943–954.
45. Lebedeva ND, Katin YA. Heats of combustion of certain monosubstituted benzenes. *Russ J Phys Chem*. 1972;46:1.
46. Weinhold F, Landis CR. Natural bond orbitals and extensions of localized bonding concepts. *Chem Educ Res Pract*. 2001;2:91–104.
47. *NBO 6.0*, Glendening ED, Badenhoop JK, Reed AE, Carpenter JE, Bohmann JA, Morales CM, Landis CR, Weinhold F, Theoretical Chemistry Institute: University of Wisconsin, Madison, 2013.
48. IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). Available at: http://goldbook.iupac.org.

*Manuscript received May 19, 2015, and revision received July 17, 2015.*