Fast correction of errors in the DFT-calculated energies of gaseous nitrogen-containing species

Ricardo Urrego-Ortiz,[a] Santiago Builes,* [a] and Federico Calle-Vallejo*[b]

[a] R. Urrego-Ortiz, Dr. S. Builes
Departamento de Ingeniería de Procesos, Universidad EAFIT, Carrera 49 No 7 sur – 50, 050022, Medellín, Colombia
E-mail: sbuiles@eafit.edu.co

[b] Dr. F. Calle-Vallejo
Departament de Ciència de Materials i Química Física & Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain
E-mail: f.calle.vallejo@ub.edu

Supporting information for this article is given via a link at the end of the document.

Abstract: Modeling adsorption phenomena on surfaces by DFT calculations often involves substantial errors, resulting in inaccurate predictions of catalytic activities. Such errors partly stem from the inaccurate description of the energetics of free molecules. Herein, we use a semiempirical group-additivity method to correct the DFT-calculated heats of formation of 106 nitrogen-containing gaseous compounds belonging to 15 different chemical families. PBE, PW91, RPBE and BEEF-vdW initially yield mean absolute errors (MAEs) with respect to experiments in the range of 0.32-0.75 eV. After correcting the systematic errors, the overall MAEs decrease to ~0.05 eV. Additionally, upon applying the corrections to three types of reaction enthalpies, the resulting MAEs are below 0.10 eV. These functional-group corrections can be used in (electro)catalysis to correct the gas-phase references necessary to evaluate equilibrium potentials and adsorption energies, predict error cancellation, and assess conflicting experimental data.

Introduction

The heat of formation is among the most important thermodynamic properties of compounds. It can be used to calculate reaction enthalpies and free energies (when combined with entropy values), which are paramount in a wide variety of applications in chemistry and chemical engineering such as energy storage,[1,2] chemical reaction engineering,[3–5] process system design,[6] combustion,[7] electrocatalysis,[8–10] and thermochemical and electrochemical stability assessment.[10–16] However, experimental heats of formation are unavailable for a vast number of compounds.[17] In such cases, computational methods provide a means to calculate those and related thermochemical properties.[18–22]

Density functional theory (DFT) is commonly used for predicting heats of formation since it allows ab-initio calculation of ground-state properties of molecules and materials.[11,23,24] The choice of the DFT exchange-correlation functional depends on the specific problem under study. For instance, exchange-correlation functionals based on the generalized gradient approximation (GGAs) usually provide a fair description of metals, while meta-GGAs and hybrid functionals are used to accurately describe molecules and solids with localized electrons.[25–27]

GGAs are known to present deviations with respect to experimental energetics of some diatomic species such as O_2.[11,12,26,28–31] In fact, numerous works have reported errors when GGA functionals are used to predict gas-phase energetics.[32–37] Despite large individual errors, GGAs may as well give fair predictions of reaction enthalpies depending on the similarity of the molecules involved. In other words, GGA functionals may benefit from error cancellation when dealing with structurally similar compounds.[13,38,39] For instance, systematic errors can be introduced by certain chemical structures, such as functional groups,[32–34] and thus GGA-based thermochemistry predictions can be closer to their experimental values when akin compounds appear on opposite sides of chemical reactions.[11,13,32]

However, it is unadvisable to rely on error cancellation when studying a series of chemical reactions. Heterogenous catalysis and surface science are specific areas where it is vital to detect and correct inaccuracies in the description of free molecules, because gas/liquid/solid interfaces have to be described at the same level of theory. If the gas phase is not well described in every step of a reaction pathway, considerable errors in the predictions may arise. An affordable approach to correct these deviations involves the use of semiempirical corrections. For instance, Peterson et al.[34] studied reaction energies of different molecules containing the OCO backbone in their structure, such as CO_2 and HCOOH. Through a minimization of the dataset's mean absolute error (MAE) with respect to experiments, an OCO backbone error of -0.45 eV was found for the RPBE exchange-correlation functional. Similarly, Studt et al.[40] determined an H_2 error of -0.09 eV for the BEEF-vdW exchange-correlation functional. Christensen et al.[33] proposed a functional analysis to identify the chemical structures that required an energy correction. They concluded that the main source of error in CO_2 reactions was associated with C=O bonds and not the OCO backbone. Guthrie[41] suggested different alternatives to calculate accurate heats of formation using DFT by following the group-additivity scheme of Benson and Buss.[42] Along those lines, Granda-Marulanda et al.[32] devised a semiempirical method to correct the formation energies of gas-phase molecules belonging to the carbon cycle by pinpointing and correcting systematic errors introduced by their functional groups.

Herein, by means of a group-additivity method, we identify and correct DFT errors in common GGAs for the calculation of heats of formation of carbon- and nitrogen-containing compounds. A dataset containing 106 gaseous molecules is used to obtain the corrections and their suitability is assessed by analyzing three different reactions. In general, the presence of nitrogen in organic compounds causes large, negative errors, and only when all corrections are simultaneously applied do the mean and maximum errors go down. We also show how our approach can aid in (i) predicting when error cancellation is significant, (ii) determining the quality of

experimental measurements when only scarce, contradicting data are available, and (iii) assessing more accurate adsorption energies and catalytic pathways.

Methodology

Computational

All calculations were performed using the VASP code[43] and four different GGAs, namely, PBE,[44] PW91,[45] RPBE,[46] and BEEF-vdW.[47] The effect of D3 dispersion corrections with and without Becke-Johnson damping[48,49] on the predictions of PBE and RPBE is discussed in section S8. The chemical families studied and the number of members in each family are provided in

Table 1 and Figure 1. Table S2 contains the list of the 106 gaseous molecules analyzed.

The DFT-calculated, ground-state energy (E_{DFT}) and the zero-point energy corrections (ZPE) were calculated for each compound. The heats of formation from the elements in their standard states were approximated as:

$$\Delta_f H^o{}_{DFT} \approx \Delta_f E_{DFT} + \Delta_f ZPE \quad (1)$$

The ZPEs were evaluated from the vibrational frequencies obtained using the harmonic oscillator approximation. The heat capacity contribution to the formation energies, $\int C_p dT$, was not included in Equation 1 since its effect is not significant in the range of 0 to 298.15 K.[11,32,50] Experimental heats of formation ($\Delta_f H^o{}_{EXP}$) were taken from thermodynamic tables for all the compounds,[51–53] except for hydroxylamine and n,o-hydroxylamine species. For hydroxylamine, the heat of formation recommended by Saraf et al.[22] was considered. For n,o-hydroxylamines species the heats of formation were taken from other works.[54]

The gaseous molecules were relaxed using the conjugate gradient algorithm in cells that assured a minimum distance of 10 Å between periodic images, with volumes up to 15625 Å^3 for the largest molecules. We sampled the reciprocal space using Monkhorst–Pack grids,[55] only considering the Γ–point. The projector augmented-wave (PAW) method[56] was used to describe ion-electron interactions. The convergence criterion for the maximal forces on the atoms was 0.01 eV/ Å. The energy cutoff was 400 eV, which guaranteed converged reaction energies (section S5.1) and enables one-to-one comparisons to previous works.[32] Gaussian smearing with an electronic temperature of 0.001 eV was used. In all cases, the energies were extrapolated to 0 K. Spin unrestricted calculations were performed only for NO.

Evaluating the errors and correcting the heats of formation

The dataset was used to determine the errors in the DFT-calculated heats of formation ($\Delta_f H^o{}_{DFT}$) of molecules with at least one of the following atoms in their structure: C, N, O or H. A general formation reaction for the molecules in the dataset is:

$$\alpha_1 C + \alpha_2 N_2 + \alpha_3 O_2 + \alpha_4 H_2 \rightarrow C_{\alpha_1} N_{2\alpha_2} O_{2\alpha_3} H_{2\alpha_4} \quad (2)$$

where α_i are stoichiometric coefficients. For instance, the formation of formamide (CH_3NO) using Equation 2 is: $C + \frac{1}{2}N_2 + \frac{1}{2}O_2 + \frac{3}{2}H_2 \rightarrow CH_3NO$. Regarding the reactants in Equation 2, the following considerations were made: the standard state of C(s) was modelled as graphene and not graphite. This approximation is supported by the fact that the interlayer cohesive energy of graphite is relatively low, as it stems from weak van der Waals interactions.[57] Reported values are in the range of 0.031–0.064 eV/atom[57–62] which, as will be shown later, are comparable to the mean errors in this study.[32] The converged C-C bond length in graphene using PBE and RPBE is 1.43 Å, and 1.42 Å using PW91 and BEEF-vdW, close to the experimental value (1.42 Å).[63] $E_{O2(g)}$ was calculated using the semiempirical approach reported in previous works (see section S1),[13] since its poor description by most DFT exchange-correlation functionals is well known.[26] H_2(g) is generally well described and only a small correction has been proposed before for BEEF-vdW.[40,64]

Equation 2 was used to calculate the heat of formation of each compound. The predicted heat of formation ($\Delta_f H^o{}_{DFT}$) without any corrections to the functional groups, was compared with the experimental heat of formation ($\Delta_f H^o{}_{EXP}$), their difference being the total error (ε_T):

$$\varepsilon_T = \Delta_f H^o{}_{DFT} - \Delta_f H^o{}_{EXP} \quad (3)$$

In line with previous works,[32] we consider that ε_T can be expressed as the sum of the errors of the reactants and products in the formation reaction. Thus, for a given reaction:

$$\varepsilon_T = \sum \varepsilon_P - \sum \varepsilon_R \quad (4)$$

where the sums collect all the errors associated to the constituent elements (ε_R) and the functional groups within the molecules (ε_P) in Equation 2, considering their stoichiometric coefficients. Assuming that the total error of a molecule depends on its functional groups is in the spirit of group additivity, which is widely used for the calculation of thermochemical properties, such as enthalpies.[42] Since C is in the solid state and O_2(g) is conveniently corrected based on H_2(g) and H_2O(g), the main contributor to the errors in the reactants in Equation 2 ($\sum \varepsilon_R$) is N_2(g).

Our analysis is based on the formation enthalpy, unlike previous works based on the Gibbs energy of formation ($\Delta_f G^o$).[32] However, the two approaches are interchangeable, since the entropies used in the computation of the Gibbs energies are usually taken from experimental tables.[32,51–53] The advantage of using the formation enthalpy instead of the free energy is the increased availability of experimental data for the former compared to the latter.

As said before, ε_R in Equation 4 is mostly related to N_2(g), and ε_{N_2} is calculated based on NH_3(g) formation (section S5), as the atomization energy of ammonia is typically well described by DFT at the GGA and meta-GGA levels.[26] This is presumably because there are only single N–H bonds in NH_3(g), and noticeable errors tend to appear in the DFT energetics of molecules when double and triple bonds are present.[26,65] The soundness of assuming a good description of ammonia is supported by the fact that the amine functional-group error is close to zero once N_2 is corrected (

Table 1), as NH_3 can be considered as an amine with R_i = –H (Figure 1).

To estimate the functional-group errors (ε_P), an averaging approach is used. The corrected heat of formation ($\Delta_f H^o{}_{CORR}$) can be calculated using Equation 5.

$$\Delta_f H^o{}_{CORR} = \Delta_f H^o{}_{DFT} - \sum \varepsilon_P + \sum \varepsilon_R \quad (5)$$

Thus, a molecule is regarded as a collection of different functional groups, each of them contributing to the total error with a specific functional-group error. The functional groups considered here are schematized in Figure 1. In general, functional groups are established from an organic chemistry perspective. When ambiguities arise, the precedence of a chemical group over another is dictated by its largest molecular weight. For instance, hydrazine (N_2H_4) and the compounds derived from it are not taken as amines with two -NR^i groups but rather as an independent family (see Figure 1). Similarly, we do not consider o-methylhydroxylamine as an ether group (-CO- with a molecular weight of 28 g/mol) bonded to an amino group (-NH_2 with molecular weight of 16 g/mol), but rather a hydroxylamine group (-ON- with a molecular weight of 30 g/mol) bonded to a methyl (-CH_3) group.

If a molecule has a unique combination of functional groups, such that no other compounds with the same chemical structure are in the dataset, its errors are assumed to be introduced by the molecule as a whole. Typical examples are CO_2 and CO, which possess well-known particular errors: the OCO-backbone and C=O bond errors, respectively.[33] Among the molecules studied here, nitric oxide (NO) belongs to this category.

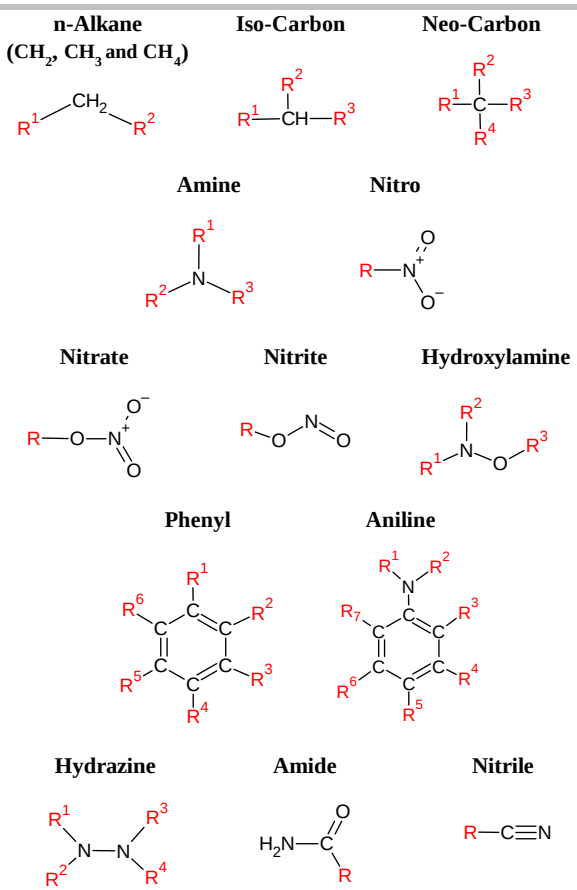

Figure 1. Functional groups considered in this work. In red, generic substituents R_i. A complete list of the compounds in this study can be found in Table S2.

Validating the corrections using reaction enthalpies

The DFT-calculated reaction enthalpy ($\Delta_r H^o{}_{DFT}$) is defined as:

$$\Delta_r H^o{}_{DFT} = \sum \Delta_f H^o{}_{DFT,P} - \sum \Delta_f H^o{}_{DFT,R} \quad (6)$$

where $\sum \Delta_f H^o{}_{DFT,P}$ collects the heats of formation of all products and $\sum \Delta_f H^o{}_{DFT,R}$ collects the heats of formation of all reactants. The corrected reaction enthalpy ($\Delta_r H^o{}_{CORR}$) is calculated by correcting the heats of formation of reactants and products, as shown in the Equation 7.

$$\Delta_r H^o{}_{CORR} = \sum \Delta_f H^o{}_{CORR,P} - \sum \Delta_f H^o{}_{CORR,R} \quad (7)$$

Results

Table *1* contains the functional-group errors (ε_P) for four exchange-correlation functionals. The calculated and experimental heats of formation for each molecule in the dataset are in Table S3. In the following, we will exemplify the error detection and quantification of the DFT-calculated heats of formation for the $-NO_3$ family using the values for PBE from Tables 1 and S3. The family includes nitric acid (HNO_3), methyl nitrate (CH_3NO_3), ethyl nitrate ($C_2H_5NO_3$), propyl nitrate ($C_3H_7NO_3$), isopropyl nitrate ($C_3H_7NO_3$) and 1, 2, 3-propanetriol trinitrate ($C_3H_5N_3O_9$). First, ε_R is determined: following Equation 2, the N_2(g) error is the only one included in the reactants: $\varepsilon_R = \frac{\alpha_2}{2} \cdot \varepsilon_{N2}$. In turn, ε_{N2}, is calculated using Equation S5. For PBE, $\varepsilon_{N2} = 0.34$ eV (

Table *1* and section S5).

The next step is the evaluation of the errors inherited from the functional groups. Nitrates have the general chemical structure R–NO_3, where R can be any functional group, for instance, an alkyl group. Therefore, ε_P for the nitrate family is:

$$\varepsilon_P = n_{NO_3} \cdot \varepsilon_{NO_3} + n_{n-alkane} \cdot \varepsilon_{n-alkane} + n_{IC} \cdot \varepsilon_{IC} + n_{NC} \cdot \varepsilon_{NC} \quad (8)$$

where n_i indicates the number of instances that the functional group i appears in the molecule, and $\varepsilon_{n-alkane}$, ε_{IC} and ε_{NC} are the errors associated to n-alkane chains, iso-Carbon and neo-Carbon groups, respectively. The term n-alkane gathers C atoms bound to several hydrogen atoms, namely, $-CH_2-$, CH_3- and CH_4. Iso-Carbon refers to an sp^3 carbon atom with only one hydrogen atom among its substituents, and neo-Carbon refers to an sp^3 carbon atom that is not bonded to any hydrogen atom.

Once the total ε_P of the compounds is expressed as a sum of the errors of each functional group, the contribution of the functional group characterizing the chemical family (ε_{NO_3}) is calculated for each family member. This presupposes that the errors of the alkyl groups are already known, so that for each compound the only unknown term in ε_P is ε_{NO_3}. See section S5 for a detailed explanation on how the separate errors per functional group were determined.

We note that the n-alkane correction was previously calculated from a dataset with molecules containing up to five carbon atoms,[32] and it was reported to a 10 meV precision. However, small rounding errors propagate when the corrections are applied to large compounds, such that cumulative residual errors can eventually surpass the MAE. This is important for the dataset used here since larger alkanes, up to ten carbon atoms (decane), are included. Thus, we recalculated the $-CH_x$ error and report it with an additional decimal place (1 meV precision) in

Table *1*. Error propagation might as well happen for the other functional groups, but $-CH_x$ moieties are ubiquitous in organic molecules, hence the importance of an accurate assessment of their errors.

For instance, consider the case of 1,2,3-propanetriol trinitrate ($C_3H_5N_3O_9$). With PBE, $\Delta_f H^o{}_{DFT}$ = -6.29 eV and $\Delta_f H^o{}_{EXP}$ = -2.89 eV, such that the total error is ε_T = -3.40 eV. The errors involved in the formation reaction of 1, 2, 3-propanetriol trinitrate, (3 C + 5/2 H_2 + 3/2 N_2 + 9/2 O_2 → $C_3H_5N_3O_9$) are those of N_2 (ε_{N_2} = 0.34 eV), the n-alkane group ($\varepsilon_{n-alkane}$ = 0.036 eV), the iso-Carbon group (ε_{IC} = 0.13 eV), and the nitrate group. Since 1,2,3–propanetriol trinitrate contains three N atoms (1.5 N_2 groups), two n-alkane groups, one iso-Carbon group and three nitrate groups, from Equation 8 we obtain ε_{NO_3} = -1.03 eV. Averaging over the entire family, the nitrate error is -1.00 eV (

Table *1*). All corrections in

Table *1* were obtained analogously.

Table 1. Error contributions of the functional groups (ε_P) present in the gas-phase heats of formation of the 106 compounds in Table S2 for four common xc-functionals. All values are in eV.

Functional Group	Number of molecules	PBE	PW91	RPBE	BEEF-vdW
N_2	1	0.34	0.38	-0.05	-0.32
NO	1	-0.07	0.05	-0.42	-0.58
CO	1	0.24	0.26	-0.10	-0.18
n-Alkane[a]	9	0.036	0.015	0.086	0.203
Iso-Carbon (IC)	7	0.13	0.12	0.25	0.20
Neo-Carbon (NC)	3	0.22	0.21	0.38	0.27
Amines	22	0.00	0.00	0.04	-0.09
Nitro	8	-0.65	-0.57	-0.81	-1.08
Nitrate	6	-1.00	-0.91	-1.17	-1.45
Nitrite	8	-0.56	-0.50	-0.79	-1.02
Hydroxylamine	5	-0.16	-0.15	-0.15	-0.31
Phenyl	9	-0.06	-0.12	-0.13	0.17
Aniline	5	-0.16	-0.21	-0.16	0.05
Hydrazine	4	-0.09	-0.08	-0.05	-0.21
Amide	9	-0.17	-0.15	-0.20	-0.38
Nitrile	9	0.10	0.11	-0.17	-0.33

[a] The error of the n-alkane functional group is reported with an additional decimal place to avoid the propagation of rounding errors for molecules with numerous -CH_x units.

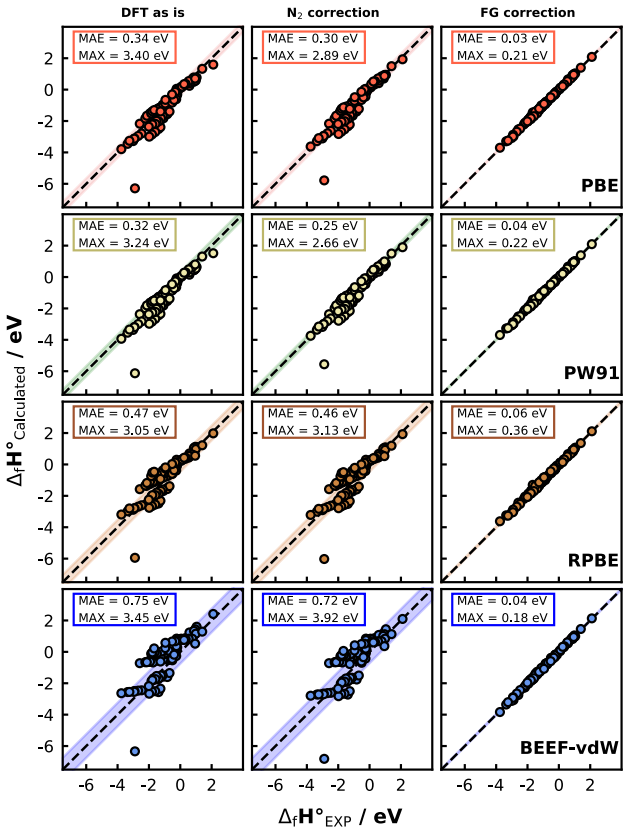

Figure 2 shows the parity plots for the experimental, DFT-calculated and corrected heats of formation. The left panels show the heats of formation before any correction (DFT as is: $\Delta_f H^o{}_{DFT}$), the middle panels show the heats of formation once N_2(g) is corrected (N_2 correction: $\Delta_f H^o{}_{DFT} + \sum \varepsilon_R$), and the right panels show the heats of formation after the functional-group corrections (FG correction: $\Delta_f H^o{}_{DFT} - \sum \varepsilon_P + \sum \varepsilon_R$). The uncorrected heats of formation contain substantial mean/maximum absolute errors (MAEs and MAXs) in the order of 0.3-0.8 and 3.0-3.5 eV, respectively. After applying the N_2 corrections (

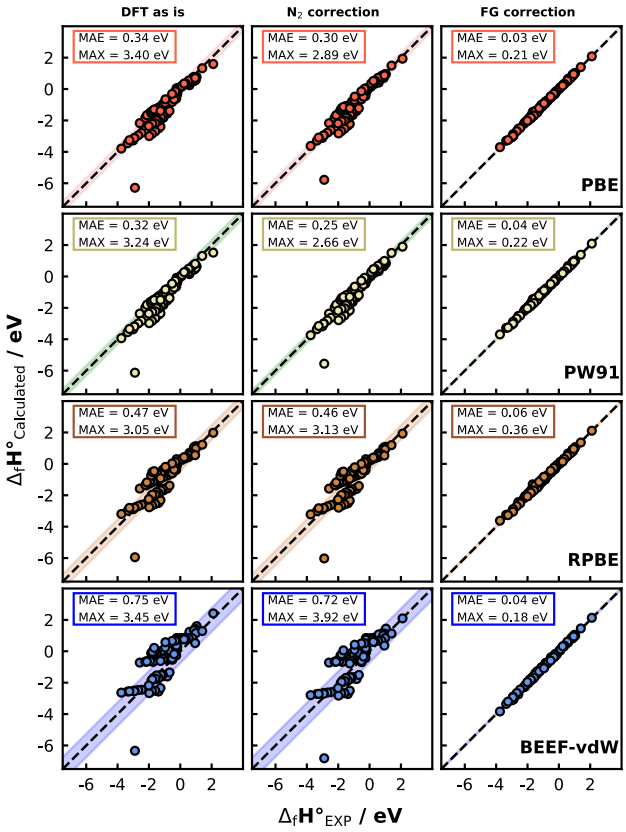

Figure 2, middle panels), the MAE of the RPBE dataset remained almost the same (0.46 eV). For the other functionals, small MAE reductions occurred. The MAX was reduced for PBE and PW91 after the N_2 correction, while showing a slight increment for RPBE (0.08 eV) and a large worsening for BEEF-vdW from 3.45 to 3.92 eV. These results suggest that the N_2(g) correction by itself does not guarantee a systematic improvement of the DFT-calculated heats of formation, and that functional-group corrections are advisable to approach chemical accuracy. The modest improvement upon the N_2(g) correction stems from the following facts:

(i) With few exceptions, the errors of the nitrogen-containing functional groups are negative and those of -CH_x groups are positive (

(ii)

(iii)

(iv) **Table *1***). Thus, for a molecule where these two errors exist, appreciable error cancellation likely takes place. Depending on the magnitude and sign of the resulting errors, the N_2(g) correction might have modest effects (e.g. RPBE) or even be detrimental for the MAX (e.g. RPBE and BEEF-vdW). See section S9 for detailed examples.

(v) $\varepsilon_R = \alpha_2 \cdot \varepsilon_{N2}$ (usually 0.17 eV for PBE, as α_2 = ½ in most cases) is markedly smaller than those of other N-containing groups, such as the nitro/nitrate groups (-0.65/-1.00 eV, see

(vi)

(vii)

(viii) **Table *1***).

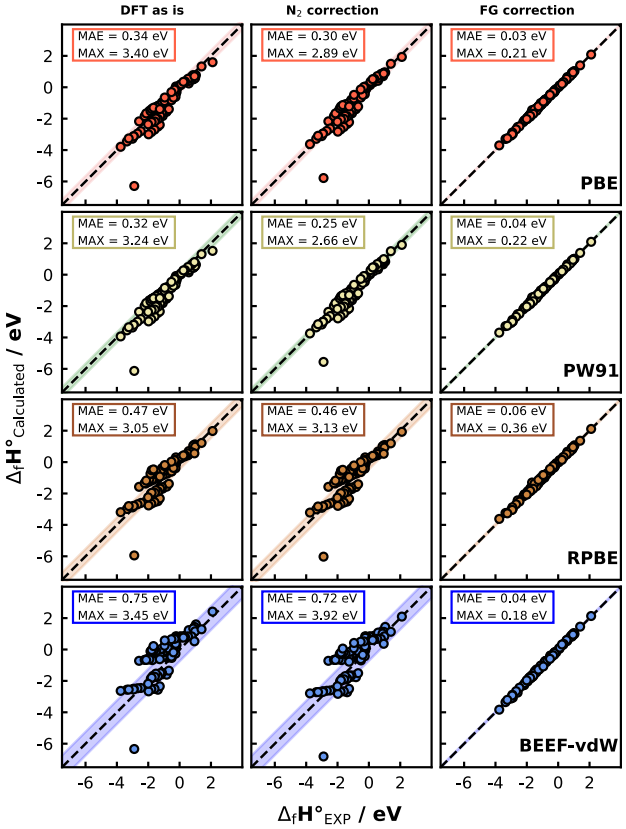

Figure 2. Parity plots for the experimental (x-axis)[22,51–54] and DFT (y-axis) heats of formation in the gas phase using PBE (red), PW91 (brown), RPBE (green) and BEEF-vdW (blue). Left: DFT-calculated heats of formation with no corrections. Center: DFT heats of formation upon correcting N_2(g). Right: heats of formation upon correcting N_2(g) and the functional group (FG) errors. The shaded area is ±MAE in each case. MAE/MAX: mean/maximum absolute errors.

We observe in

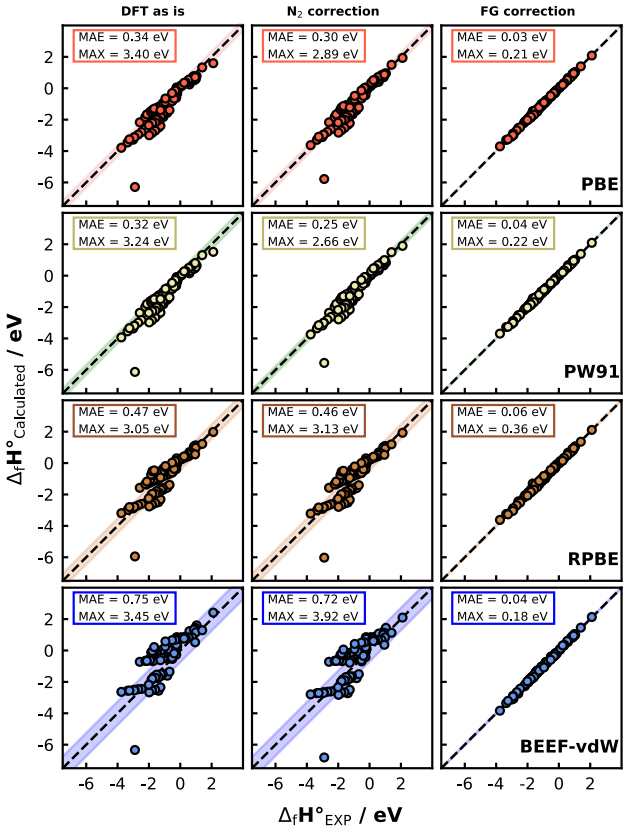

Figure 2 (right panels) that when the corrections to the functional groups were applied, the overall MAE and MAX of the dataset were lowered by one order of magnitude, and in all cases approached chemical accuracy (1 kcal/mol). Figures S1-S13 show that compensating each functional-group error contributes appreciably to lowering the MAE and MAX of the respective family of compounds. Hence, the assignment of the compounds to their respective chemical families is adequate, and the functional-group errors are systematic.

The corrections in

Table *1* are functional dependent and, thus, different values and signs are observed for the analyzed exchange-correlation functionals. For a given family, similar errors were obtained for PBE and PW91, as has been shown for other energetic calculations.[32,66,67] This reflects the fact that PBE was devised to mimic PW91's results through a simpler approach.[44] In turn, the corrections for RPBE in

Table *1* are generally halfway between those of PW91/PBE and BEEF-vdW.

Table *1* provides an easy way to predict if a given molecule can be accurately described without any corrections. For instance, hydrazine (N_2H_4) calculated with RPBE has a similar error to that of N_2. In this case, $\sum \varepsilon_P - \sum \varepsilon_R$= 0.00 eV, such that the DFT heat of formation is equal to the experimental and the corrected ones. For n-alkanes, this approach is useful. For example, for PW91 $\varepsilon_{n-alkane}$ = 0.015 eV, thus, short-chain n-alkanes (up to n-butane) are described using this functional with errors below 0.1 eV. Conversely, the use of BEEF-vdW or RPBE (with errors of 0.20 eV/CH_x and 0.09 eV/CH_x) definitely entails gas-phase corrections if accurate energetics are sought.

Because the method presented here is semiempirical, the quality and availability of the experimental data are paramount. Nevertheless, in some cases, experimental data are scarce or inconsistent, leading to ambiguous error estimations. For instance, the experimental heat of formation of hydroxylamine has been reported in previous works with differences of 0.08 eV.[54,68] This error is comparable to the maximum error (MAX) of the hydroxylamine family found for the four functionals upon the corrections (Figures 1 and S8). Thus, one may use the parity line to identify "suspicious" experimental values, corresponding to compounds that do not fall on the parity line after all necessary corrections. For example, Figure S4 (right panels) shows that for all functionals, except BEEF-vdW, a dissonant value for tripropylamine persists after all corrections are applied. This might be an indication that the experimental data have large uncertainties. In fact, the reliability of the experimental data cannot be ascertained since the source does not report one of the references used in the determination of the heat of formation.[53]

The presence of several groups within a given molecule may lead to different gas-phase corrections, which can be obtained analogously to the rest of the corrections in this study. This is illustrated for $CH_x(NO_2)_y$ (x = 1,2,3; x+y = 4) in section S7.

The corrections in

Table *1* were suitably applied to all compounds in the dataset to correct the reaction energy of Equation 9, in which the molecules in the dataset are produced from CO(g), NO(g) and H_2O(g).

$$\alpha_1 \cdot CO + 2\alpha_2 \cdot NO + \alpha_4 \cdot H_2O \rightarrow C_{\alpha_1}N_{2\alpha_2}O_{2\alpha_3}H_{2\alpha_4} + \alpha_5 \cdot O_2 \qquad (9)$$

where $C_{\alpha_1}N_{\alpha_2}O_{2\alpha_3}H_{2\alpha_4}$ refers to the molecules within the dataset (as in Equation 2), and $\alpha_5 = (\alpha_1 + 2\alpha_2 - 2\alpha_3 + \alpha_4)/2$. For instance, Equation 9 for formamide is: $CO + NO + \frac{3}{2}H_2O \rightarrow CH_3NO + \frac{5}{4}O_2$. In Table S4 we provide the experimental and calculated reaction energies that produce each compound in the dataset.

Figure ***3*** contains the parity plots of the corresponding experimental ($\Delta_r H^o{}_{EXP}$) and calculated reaction enthalpies ($\Delta_r H^o{}_{Calculated}$), see Table S4. Large errors are observed in the uncorrected reaction energies shown in the left panels of

Figure ***3***, with the MAEs and MAXs in the ranges of 1.0–1.7 and 2.5–3.9 eV, respectively. After applying the corrections to the reactants (

Figure ***3***, center panels), a MAE lowering of ~1.0 eV is achieved for RPBE and BEEF-vdW, while reductions of ~0.8 and 0.5 eV occurred for PBE and PW91, respectively. The final errors are close to chemical accuracy (

Figure ***3***, right panels). The MAEs and MAXs of the reaction enthalpies are identical to the errors of the heats of formation in the middle and right panels in Figure 2, which shows that the functional-group errors are systematic and do not propagate upon the corrections.

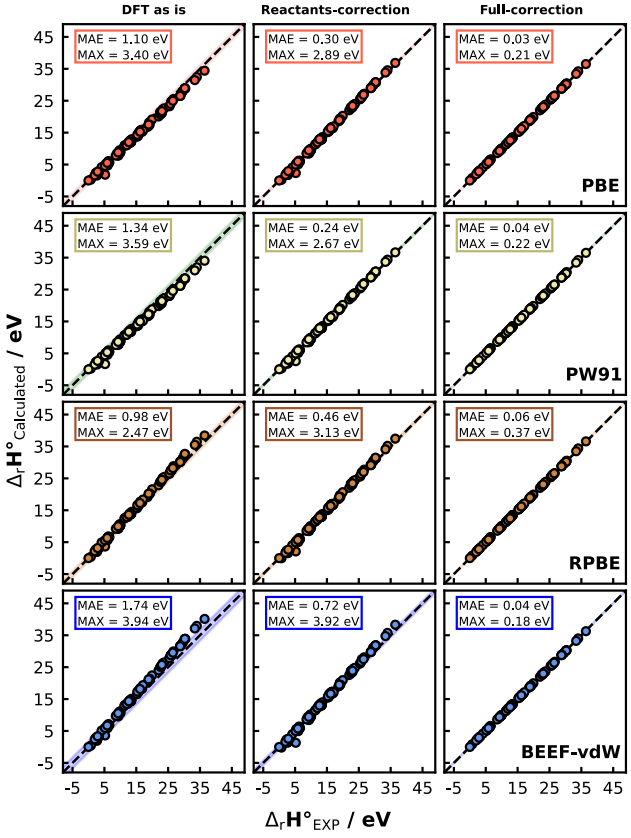

Figure 3. Parity plots for the experimental (x-axis)[22,51–54] and DFT (y-axis) reaction enthalpies of Equation 9 using PBE (red), RPBE (green), PW91 (brown), and BEEF-vdW (blue). Left: DFT-calculated reaction enthalpies with no corrections. Center: DFT reaction enthalpies upon correcting N_2(g). Right: heats of formation upon correcting N_2(g) and the functional groups. The shaded area is ±MAE in each case. MAE/MAX: mean/maximum absolute errors.

The relevance of the semiempirical method for energy calculations in chemicals reactions is highlighted by the possibility of reactions modifying the functional groups. As seen in

Table 1, the errors of N-containing compounds may exhibit large variations depending on the functional group, and it is expected that error cancelation will not forcedly apply to all reactions involving those functional groups. Thus, as a final test, we consider two reactions combining substances within the dataset, namely the reduction of nitrates to nitrites and the isomerization of nitro-compounds to nitrites (Equations 10 and 11, respectively). In Table S5 the reactants, products and substituents (R_i) are detailed.

$$R - NO_3 + H_2 \rightarrow R - ONO + H_2O \quad (10)$$

$$R - NO_2 \rightarrow R - ONO \quad (11)$$

Figure 4 shows the parity plots between the experimental ($\Delta_r H^o{}_{EXP}$) and calculated values ($\Delta_r H^o{}_{Calculated}$) for these reactions. The left panels contain the data of the reduction reaction, while the right panels contain the isomerization data. Each panel shows the uncorrected (blue circles) and corrected (orange crosses) reaction enthalpies.

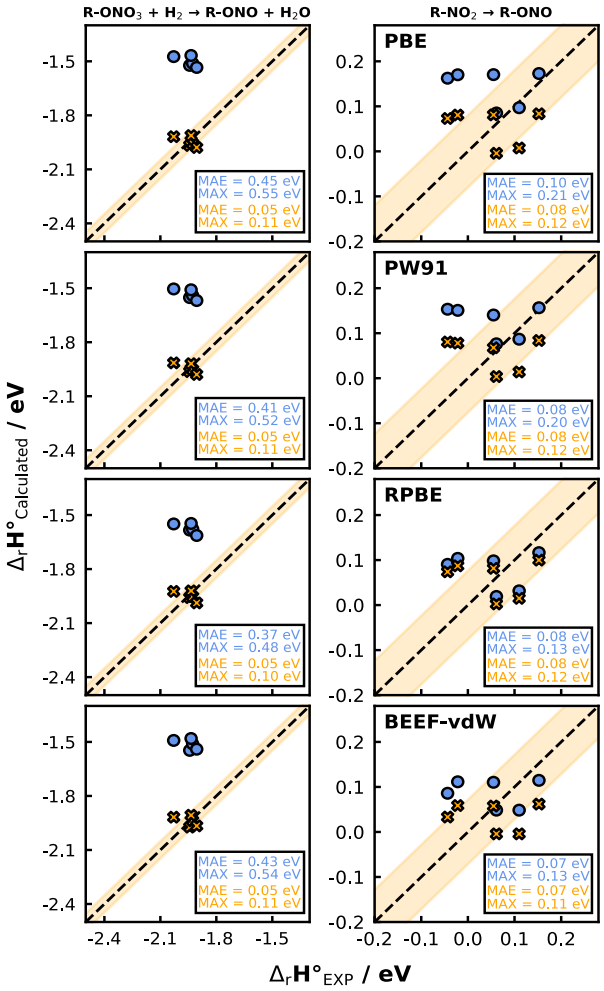

Figure 4. Parity plots for the experimental (x-axis)[22,51–54] and DFT (y-axis) reaction enthalpies for the reduction of nitrates to nitrites (left) and the isomerization of nitro-compounds to nitrites (right) using PBE, RPBE, PW91, and BEEF-vdW. The reaction enthalpies with no corrections are shown as blue circles. The corrected reaction enthalpies are shown as orange crosses. The uncorrected/corrected MAEs and MAXs are shown in blue/orange. The shaded area is ±MAE of the corrected data in each case

As in Figures 2 and 3, large errors are present in the uncorrected enthalpies of nitrate reduction (average of the MAEs ≈ 0.42 eV; MAX ≈ 0.55 eV). This is expectable, as the alkane-based errors in Equation 10 cancel out, and the total error is due to the -NO_3 and -ONO groups, which differ by ~0.42 eV (

Table *1*). Upon the corrections, there is a clear shift toward the parity line (average of the MAEs ≈ 0.05 eV; MAX ≈ 0.11 eV).

On the other hand, for the isomerization reaction, the errors before the corrections are not large (average of the MAEs ≈ 0.08 eV; MAX ≈ 0.21 eV). This is also expectable, as the corrections for the nitrite and nitro groups differ only by 0.06 eV (

Table 1). When the corrections are applied, the MAEs remain small and the MAXs decrease (average of the MAEs ≈ 0.08 eV; MAX ≈ 0.12 eV). In brief,

Figure 4 shows that when the corrected energies of the substances in the dataset are combined, the resulting errors in the reaction energies are also low, without relying on error cancellation.

Brief discussion

Our method can be used for the accurate determination of reaction energies and adsorption energies in heterogeneous catalysis, and of equilibrium potentials and adsorption energies in electrocatalysis. This is schematized in Figure 5 for the following hypothetical reaction:

$A_{(g)} + B_{(g)} \rightarrow AB_{(g)}$ (12)

The enthalpy of this gas-phase reaction can be calculated from the heats of formation of $A_{(g)}$ and $B_{(g)}$ and $AB_{(g)}$:

$\Delta_r H = \Delta_f H_{AB} - \Delta_f H_A - \Delta_f H_B$ (13)

Clearly, if there are appreciable errors in $\Delta_f H_A$, $\Delta_f H_B$, $\Delta_f H_{AB}$ calculated with DFT it is likely that the heat of reaction ($\Delta_r H$) will deviate considerably from experiments, unless there is significant error cancellation. Furthermore, the gas-phase errors also affect the adsorption energies of A, B and AB (ΔH_A^{ads}, ΔH_B^{ads}, ΔH_{AB}^{ads}), so that the analysis of catalytic pathways might be severely affected. According to Figure 5, the connection between the catalytic steps and the reaction energy is the following:

$\Delta_r H = \Delta H_A^{ads} + \Delta H_B^{ads} + \Delta H_{coupling} - \Delta H_{AB}^{ads}$ (14)

Thus, the first step toward a good description of catalytic pathways is a good estimation of the overall reaction enthalpy and its separate components. Of course, more ingredients are needed for an accurate assessment of adsorption energies in solution, namely corrections to the adsorbed state[33,69] and solvation corrections.[70–72] In fact, recent studies coupling gas-phase corrections and solvation corrections led to reaction free energies, equilibrium potentials and catalytic onset potentials in good agreement with experiments.[32,73]

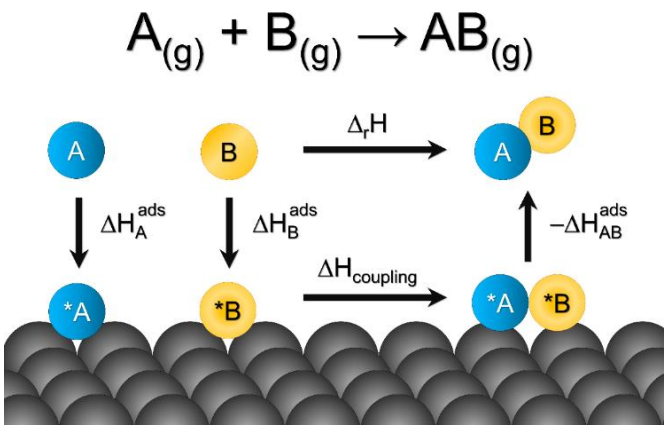

Figure 5. Hypothetical catalytic coupling of A and B to produce AB. The catalytic coupling entails the adsorption of A and B, their surface coupling, and the desorption of AB. Except for *A-*B coupling, all steps involve gas-phase species and, thus, are affected by their associated errors.

Conclusions

Here we showed that when PBE, PW91, RPBE and BEEF-vdW, are used to predict the heats of formation of N-containing gaseous compounds, large average errors are encountered. Such errors are introduced by molecular nitrogen (N_2) and an assortment of functional groups present in the compounds. Using a semiempirical method, we observed that correcting the error in N_2(g) alone does not yield a significant improvement of the dataset's DFT-calculated predicted enthalpies. Conversely, when both nitrogen and functional group corrections were implemented, the MAEs and MAXs decrease by one order of magnitude, and the resulting MAEs are close to chemical accuracy.

Our correction scheme is robust enough to correct the DFT-calculated reaction enthalpies regardless of the presence or absence of error cancellation. This was illustrated by applying the corrections to the reduction reaction of nitrates to nitrites, which decreased the MAEs from ~0.42 to ~0.05 eV. In addition, the MAEs for the isomerization of nitro-compounds to nitrites remained at ~0.08 eV after the corrections. Thus, the method allows a rapid assessment of the need for gas-phase corrections for given substances calculated with specific exchange-correlation functionals and can be used to anticipate error cancellation in (electro)chemical reactions. In addition, because the method systematically leads to lower errors with respect to experiments, it can be used to evaluate the reliability of experimental heats of formation.

In brief, our semiempirical method provides a simple and fast way to correct DFT errors in the heats of formation for a wide variety of N-containing gaseous compounds. This is useful in computational heterogenous catalysis and electrocatalysis, where an accurate assessment of equilibrium potentials and adsorption energies presupposes a good description of gaseous molecules.

Acknowledgements

This work was supported by Universidad EAFIT through project 952-000018. F.C.V. thanks the Spanish MICIUN for a Ramón y Cajal research contract (RYC-2015-18996), the RTI2018-095460-B-I00 grant and financial support through the program "Units of Excellence María de Maeztu" (grant MDM-2017–0767). We also acknowledge funding by Generalitat de Catalunya (2017SGR13). The use of supercomputing facilities at SURFsara was sponsored by NWO Physical Sciences, with financial support by NWO. The authors also acknowledge the use of supercomputing resources of the Centro de Computación Científica Apolo at Universidad EAFIT (http://www.eafit.edu.co/apolo). This research used resources of the Center for Functional Nanomaterials, which is a U.S. DOE Office of Science Facility, and the Scientific Data and Computing Center, a component of the Computational Science Initiative, at Brookhaven National Laboratory under Contract No. DE-SC0012704.

Keywords: gas-phase errors • gas-phase reactions • heats of formation • generalized gradient approximation • nitrogen cycle

[1] J. Yang, A. Sudik, C. Wolverton, D. J. Siegel, *Chem. Soc. Rev.* **2010**, *39*, 656–675.
[2] V. L. Chevrier, S. P. Ong, R. Armiento, M. K. Y. Chan, G. Ceder, *Phys. Rev. B* **2010**, *82*, 075122.
[3] F. Hess, B. M. Smarsly, H. Over, *Acc. Chem. Res.* **2020**, *53*, 380–389.
[4] M. Saab, F. Réal, M. Šulka, L. Cantrel, F. Virot, V. Vallet, *J. Chem. Phys.* **2017**, *146*, 244312.
[5] V. V. Chaban, O. V. Prezhdo, *J. Phys. Chem. Lett.* **2016**, *7*, 2622–2626.
[6] C.-C. Tsai, S.-T. Lin, *AIChE J.* **2020**, *66*, e16987.
[7] A. Osmont, L. Catoire, I. Gökalp, V. Yang, *Combust. Flame* **2007**, *151*, 262–273.
[8] J. Greeley, I. E. L. Stephens, A. S. Bondarenko, T. P. Johansson, H. A. Hansen, T. F. Jaramillo, J. Rossmeisl, I. Chorkendorff, J. K. Nørskov, *Nat. Chem.* **2009**, *1*, 552–556.
[9] F. Calle-Vallejo, O. A. Díaz-Morales, M. J. Kolb, M. T. M. Koper, *ACS Catal.* **2015**, *5*, 869–873.
[10] J. I. Martínez, H. A. Hansen, J. Rossmeisl, J. K. Nørskov, *Phys. Rev. B* **2009**, *79*, 045120.
[11] G. Hautier, S. P. Ong, A. Jain, C. J. Moore, G. Ceder, *Phys. Rev. B* **2012**, *85*, 155208.
[12] H. Smithson, C. A. Marianetti, D. Morgan, A. Van der Ven, A. Predith, G. Ceder, *Phys. Rev. B* **2002**, *66*, 144107.
[13] F. Calle-Vallejo, J. I. Martínez, J. M. García-Lastra, M. Mogensen, J. Rossmeisl, *Angew. Chem., Int. Ed.* **2010**, *49*, 7699–7701.
[14] J. Cheng, A. Navrotsky, X.-D. Zhou, H. U. Anderson, *J. Mater. Res.* **2005**, *20*, 191–200.
[15] A. S. Raman, R. Patel, A. Vojvodic, *Faraday Discuss.* **2020**, DOI 10.1039/C9FD00146H.
[16] A. Zagalskaya, V. Alexandrov, *ACS Catal.* **2020**, *10*, 3650–3657.
[17] S. V. Meschel, X. Q. Chen, O. J. Kleppa, P. Nash, *CALPHAD: Comput. Coupling Phase Diagrams Thermochem.* **2009**, *33*, 55–62.
[18] D. Jha, K. Choudhary, F. Tavazza, W. Liao, A. Choudhary, C. Campbell, A. Agrawal, *Nat. Commun.* **2019**, *10*, 5316.
[19] J. Hafner, C. Wolverton, G. Ceder, *MRS Bull.* **2006**, *31*, 659–668.
[20] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, C. Wolverton, *npj Comput. Mater.* **2015**, *1*, 15010.
[21] P. M. Mayer, C. J. Parkinson, D. M. Smith, L. Radom, *J. Chem. Phys.* **1998**, *108*, 604–615.
[22] S. R. Saraf, W. J. Rogers, M. S. Mannan, M. B. Hall, L. M. Thomson, *J. Phys. Chem. A* **2003**, *107*, 1077–1081.
[23] A. Jain, Y. Shin, K. A. Persson, *Nat. Rev. Mater.* **2016**, *1*, 15004.
[24] O. V. Kharissova, B. I. Kharisov, L. T. González, *J. Mater. Res.* **2020**, *35*, 1424–1438.
[25] P. Janthon, S. (Andy) Luo, S. M. Kozlov, F. Viñes, J. Limtrakul, D. G. Truhlar, F. Illas, *J. Chem. Theory Comput.* **2014**, *10*, 3832–3839.
[26] S. Kurth, J. P. Perdew, P. Blaha, *Int. J. Quantum Chem.* **1999**, *75*, 889–909.
[27] J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, *Phys. Rev. Lett.* **2003**, *91*, 146401.
[28] R. O. Jones, O. Gunnarsson, *Rev. Mod. Phys.* **1989**, *61*, 689–746.
[29] L. A. Curtiss, K. Raghavachari, P. C. Redfern, J. A. Pople, *J. Chem. Phys.* **1997**, *106*, 1063–1079.
[30] S. Grindy, B. Meredig, S. Kirklin, J. E. Saal, C. Wolverton, *Phys. Rev. B* **2013**, *87*, 075150.

[31] D. C. Patton, D. V. Porezag, M. R. Pederson, *Phys. Rev. B* **1997**, *55*, 7454–7459.
[32] L. P. Granda-Marulanda, A. Rendón-Calle, S. Builes, F. Illas, M. T. M. Koper, F. Calle-Vallejo, *ACS Catal.* **2020**, *10*, 6900–6907.
[33] R. Christensen, H. A. Hansen, T. Vegge, *Catal. Sci. Technol.* **2015**, *5*, 4946–4949.
[34] A. A. Peterson, F. Abild-Pedersen, F. Studt, J. Rossmeisl, J. K. Nørskov, *Energy Environ. Sci.* **2010**, *3*, 1311–1315.
[35] F. Calle-Vallejo, M. Huang, J. B. Henry, M. T. M. Koper, A. S. Bandarenka, *Phys. Chem. Chem. Phys.* **2013**, *15*, 3196–3202.
[36] X. Huang, S. E. Mason, *Surf. Sci.* **2014**, *621*, 23–30.
[37] O. V. Gritsenko, Ł. M. Mentel, E. J. Baerends, *J. Chem. Phys.* **2016**, *144*, 204114.
[38] V. Stevanović, S. Lany, X. Zhang, A. Zunger, *Phys. Rev. B* **2012**, *85*, 115104.
[39] L. Wang, T. Maxisch, G. Ceder, *Phys. Rev. B* **2006**, *73*, 195107.
[40] F. Studt, F. Abild-Pedersen, J. B. Varley, J. K. Nørskov, *Catal. Lett.* **2013**, *143*, 71–73.
[41] J. P. Guthrie, *J. Phys. Chem. A* **2001**, *105*, 9196–9202.
[42] S. W. Benson, J. H. Buss, *J. Chem. Phys.* **1958**, *29*, 546–572.
[43] G. Kresse, J. Furthmüller, *Phys. Rev. B* **1996**, *54*, 11169–11186.
[44] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
[45] J. P. Perdew, K. Burke, Y. Wang, *Phys. Rev. B* **1996**, *54*, 16533–16539.
[46] B. Hammer, L. B. Hansen, J. K. Nørskov, *Phys. Rev. B* **1999**, *59*, 7413–7421.
[47] J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard, K. W. Jacobsen, *Phys. Rev. B* **2012**, *85*, 235149.
[48] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[49] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.
[50] C. J. Bartel, A. W. Weimer, S. Lany, C. B. Musgrave, A. M. Holder, *npj Comput. Mater.* **2019**, *5*, 4.
[51] D. R. Lide, *CRC Handbook of Chemistry and Physics, 90th Edition (CD-ROM Version 2010)*, CRC Press/Taylor And Francis, Boca Raton, FL, **2010**.
[52] J. B. Pedley, R. D. Naylor, S. P. Kirby, *Thermochemical Data of Organic Compounds*, Springer Netherlands, Dordrecht, **1986**.
[53] P. J. Linstrom, W. G. Mallard, Eds., *NIST Chemistry WebBook, NIST Standard Reference Database 69*, National Institute Of Standards And Technology, Gaithersburg MD, **1997**.
[54] D. S. Shaikhlislamov, M. R. Talipov, S. L. Khursan, *Russ. J. Phys. Chem. A* **2007**, *81*, 235–240.
[55] H. J. Monkhorst, J. D. Pack, *Phys. Rev. B* **1976**, *13*, 5188–5192.
[56] G. Kresse, D. Joubert, *Phys. Rev. B* **1999**, *59*, 1758–1775.
[57] W. Wang, S. Dai, X. Li, J. Yang, D. J. Srolovitz, Q. Zheng, *Nat. Commun.* **2015**, *6*, 7853.
[58] L. A. Girifalco, R. A. Lad, *J. Chem. Phys.* **1956**, *25*, 693–697.
[59] R. Zacharia, H. Ulbricht, T. Hertel, *Phys. Rev. B* **2004**, *69*, 155406.
[60] L. X. Benedict, N. G. Chopra, M. L. Cohen, A. Zettl, S. G. Louie, V. H. Crespi, *Chem. Phys. Lett.* **1998**, *286*, 490–496.
[61] M. Xia, C. Liang, Z. Cheng, R. Hu, S. Liu, *Phys. Chem. Chem. Phys.* **2019**, *21*, 1217–1223.
[62] Z. Liu, J. Z. Liu, Y. Cheng, Z. Li, L. Wang, Q. Zheng, *Phys. Rev. B* **2012**, *85*, 205418.
[63] D. C. Elias, R. R. Nair, T. M. G. Mohiuddin, S. V. Morozov, P. Blake, M. P. Halsall, A. C. Ferrari, D. W. Boukhvalov, M. I. Katsnelson, A. K. Geim, K. S. Novoselov, *Science* **2009**, *323*, 610–613.
[64] F. Studt, M. Behrens, E. L. Kunkes, N. Thomas, S. Zander, A. Tarasov, J. Schumann, E. Frei, J. B. Varley, F. Abild-Pedersen, J. K. Nørskov, R. Schlögl, *ChemCatChem* **2015**, *7*, 1105–1111.
[65] K. P. Kepp, *Commun. Chem.* **2018**, *1*, 63.
[66] B.-T. Teng, X.-D. Wen, M. Fan, F.-M. Wu, Y. Zhang, *Phys. Chem. Chem. Phys.* **2014**, *16*, 18563–18569.
[67] L. P. Granda-Marulanda, S. Builes, M. T. M. Koper, F. Calle-Vallejo, *ChemPhysChem* **2019**, *20*, 2968–2972.
[68] L. V. Gurvich, I. V. Veits, C. B. Alcock, *Thermodynamic Properties of Individual Substances. Volume 1 - Elements O, H/D, T/, F, Cl, Br, I, He, Ne, Ar, Kr, Xe, Rn, S, N, P, and Their Compounds. Part 1 - Methods and Computation. Part 2 - Tables (4th Revised and Enlarged Edition)*, Hemisphere Pub. Corp., New York, **1989**.
[69] R. Christensen, H. A. Hansen, C. F. Dickens, J. K. Nørskov, T. Vegge, *J. Phys. Chem. C* **2016**, *120*, 24910–24916.
[70] Y. Basdogan, A. M. Maldonado, J. A. Keith, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *10*, e1446.
[71] V. Tripkovic, *Phys. Chem. Chem. Phys.* **2017**, *19*, 29381–29388.
[72] S. Sakong, A. Groß, *ACS Catal.* **2016**, *6*, 5575–5586.
[73] A. Rendón-Calle, S. Builes, F. Calle-Vallejo, *Appl. Catal., B* **2020**, *276*, 119147.