Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Unifying thermochemistry concepts in computational heterogeneous catalysis†

Bjarne Kreitz,*[a] Gabriel S. Gusmão,[b] Dingqi Nai,[b] Sushree Jagriti Sahoo,[b], Andrew A. Peterson,[a] David H. Bross,[c], C. Franklin Goldsmith*,[a] and Andrew J. Medford*[b]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Thermophysical properties of adsorbates and gas-phase species define the free energy landscape of heterogeneously catalyzed processes and are pivotal for an atomistic understanding of the catalyst performance. These thermophysical properties, such as the free energy or the enthalpy, are typically derived from density functional theory (DFT) calculations. Enthalpies are species-interdependent properties that are only meaningful when referenced to other species. The widespread use of DFT has led to a proliferation of new energetic data in the literature and databases. However, there is a lack of consistency in how DFT data is referenced and how the associated enthalpies or free energies are stored and reported, leading to challenges in reproducing or utilizing the results of prior work. Additionally, DFT suffers from exchange-correlation errors that often require corrections to align the data with other global thermochemical networks, which are not always clearly documented or explained. In this review, we introduce a set of consistent terminology and definitions, review existing approaches, and unify the techniques using the framework of linear algebra. This set of terminology and tools facilitates the correction and alignment of energies between different data formats and sources, promoting the sharing and reuse of *ab initio* data. Standardization of thermochemistry concepts in computational heterogeneous catalysis reduces computational cost and enhances fundamental understanding of catalytic processes, which will accelerate the computational design of optimally performing catalysts.

1 Introduction

Density functional theory (DFT) has emerged as a crucial tool in computational chemistry over the last decades, revolutionizing the field of heterogeneous catalysis by providing an effective approach to predict energetic properties[1–6]. This quantum-mechanical (QM) modeling method offers insight into surface chemistry and catalysis[7,8]. First-principles-based multiscale modeling provides the bridge between experiments and theory that enables optimizations of catalysts and reactors[9–11]. The DFT developments combined with advances in computational power have led to a surge of large-scale open-source databases, such as the Materials Project[12], the Open Catalyst Project (OCP)[13–15], the Novel Materials Discovery Laboratory (NoMaD)[16–18], Catalysis Hub[19], Automatic FLOW for Materials Discovery (AFLOW)[20], CatApp[21], and the Open Quantum Material Database (OQMD)[22,23]. These databases have enabled the proliferation of energetic data, which has allowed a deeper understanding of catalytic processes at the atomic level.

Although these databases provide a vast amount of computed information, the absence of standardized data formats and naming conventions poses a challenge for researchers. The databases typically do not contain the raw DFT energies of the structures but rather the derived quantities such as thermophysical properties and activation barriers. In principle, these derived quantities are more general and transferable than raw DFT energies. However, the processes used to derive them and the terminology used to describe them can lead to inconsistencies between different studies[24]. This challenge is further compounded by the variety of DFT codes and the different levels of theory used to generate the data[25,26]. Due to computational cost, most large-scale databases are based on DFT calculations with the generalized gradient (GGA) approximation. Such DFT codes include The Vienna Ab initio Simulation Package (VASP)[27,28], Quantum ESPRESSO[29,30], NWChem[31], GPAW[32], Jaguar[33], and SPARC[34],

[a] School of Engineering, Brown University, Providence, Rhode Island 02912, United States; E-mail: bjarne_kreitz@brown.edu, franklin_goldsmith@brown.edu
[b] School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States; E-mail: ajm@gatech.edu
[c] Chemical Sciences and Engineering Division, Argonne National Laboratory, Lemont, Illinois 60439, United States

† Electronic Supplementary Information (ESI) available: DFT data, additional mathematical derivations, and Jupyter notebooks. See DOI: 00.0000/00000000.

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

each of which has its own features, functionalities, and specializations. The Atomic Simulation Environment (ASE)[35] or Python Materials Genomics (pymatgen)[36] have helped standardize the interfaces to these codes, but numerical differences persist due to differences in exchange-correlation (xc) functionals, settings (e.g. plane-wave cutoff energy, k-point sampling) and pseudopotentials[25]. For metal oxides and magnetic systems, this is further complicated by the spin state optimization and Hubbard correction (DFT+U)[37,38], while electrochemical systems are affected by the potential and solvent model[39–43]. Advanced techniques such as quantum Monte Carlo, the random phase approximation, or time-dependent DFT can improve physical accuracy but also introduce additional sources of numerical error and ambiguity[44–49].

Consequently, it is challenging to efficiently access, compare, integrate, and analyze this data because different studies and databases often present their findings in different formats[25,50–52]. Web applications for surface chemistry such as CatApp[21] and Catalysis-Hub[19] aim to provide systematic data for activation energies of elementary surface reactions and binding energies from a variety of studies. Yet, the broader challenge of standardizing data across various platforms and integrating existing gas-phase databases remains. The rise of machine learning (ML) in heterogeneous catalysis research has been facilitated by the availability of these databases and the development of cutting-edge approaches to handle the data, leading to the discovery of optimized catalysts[53–56]. However, the inconsistencies between datasets impedes the seamless data integration and (re)use, which hinders the development of better ML models[24,50,57–59]. These discrepancies underscore the need for standardized practices and benchmarks in DFT calculations to ensure consistency and reproducibility across studies[60,61]. Due to all of the reasons above, it is common for every investigator of a catalytic reaction to perform electronic structure calculations for all species with a consistent set of DFT settings to obtain meaningful results. Given the millions of calculations present in emerging DFT energy databases[14,15], it is imperative that researchers find ways to leverage and reuse existing DFT data whenever possible.

Beyond the challenges posed by non-standardized data formats and different DFT methods, another critical issue that must be addressed is the potential inconsistencies between experimental and *ab initio* data. Understanding catalytic processes relies on microkinetic modeling based on DFT-derived energetics[9,10,62]. However, the uncertainties in the energetics lead to discrepancies between theory and experiment, which impact the applicability and reliability of the multiscale model[9,63,64]. Discrepancies can arise from various factors, such as incorrect active site models[65], missing pathways in the mechanism, inadequate consideration of surface environment under reaction conditions[66], and intrinsic errors in the DFT xc functional[67]. DFT suffers from relatively large physical errors arising primarily from the xc approximation, which yields adsorption energies errors of $\approx 20\,\mathrm{kJ\,mol^{-1}}$ [67]. While DFT has been instrumental in studying catalytic materials, intrinsic parametric errors translate to uncertainties in reaction rates of multiple orders of magnitude in multiscale models [68–72]. Error cancellation improves the reliability of trends calculated with DFT, but the errors are still too large for quantitative predictions of reaction rates for individual materials with first-principles-based kinetic models in most cases.

The relatively mature field of gas-phase chemistry has made significant progress in establishing common references and protocols for aligning thermochemical data from various sources. Together with highly accurate electronic structure methods, this has led to the development of databases with highly accurate thermochemical properties, *i.e.* enthalpies of formation ($\Delta_f H$) such as the NIST Chemistry Webbook[73], JANAF[74], or the Active Thermochemical Tables (ATcT)[75,76]. The accuracy and consistency of the gas-phase thermochemical data are the cornerstone of quantitative predictions of reaction rates, which bridge the gap between theoretical predictions and experimental observations. When a gas-phase microkinetic model is developed, it is possible to simply use tabulated values and append only the thermophysical properties of missing species.

Drawing inspiration from the approaches the gas-phase community uses for aligning thermochemical states offers a promising route for standardizing approaches in the heterogeneous catalysis community. However, heterogeneous catalysis has distinct challenges, as outlined above. Errors in electronic structure methods for heterogeneous systems at the GGA-level are larger[77], and there is no clear hierarchy of methods that can be universally applied across different systems[60,78,79]. Furthermore, the energies of adsorbates are intricately linked to the surface chemistry, making them highly dependent on the specific catalyst surface, adsorbate coverage, and reaction conditions[80–83]. Given the intricate nature of heterogeneous catalysis, a more nuanced strategy is necessary to standardize methodologies, emphasizing the urgency and importance of this task.

In this review, we first define a consistent notation of thermochemical quantities that the heterogeneous catalysis community can use in the future and provide a bridge to the gas-phase thermochemical frameworks. The structure of this review ensures that the methods of deriving thermophysical properties from QM calculations are integrated into the general thermochemistry framework to provide a clear learning pathway. We review the available approaches to derive thermochemical properties of adsorbates from *ab initio* data using a standardized notation that can be condensed into a simple set of linear algebra equations. These linear algebra tools provide a straightforward method for combining data from numerous sources, integrating experiments with *ab initio* data, converting between different formats, identifying and minimizing inconsistencies, and guaranteeing that the combined dataset is complete and consistent. The various approaches are used to calculate the thermophysical parameters for a case study to showcase differences and similarities. We restrict the discussion to heterogeneous catalysis for gas/solid systems for simplicity, although extensions to solution-phase chemistry[39] and electrochemistry[84] are expected to be relatively straightforward. Our aim is not to promote any single method but rather to demonstrate the differences, benefits, and drawbacks of available techniques while providing an approachable introduction to a topic often perplexing to novice researchers in the field. We hope that this review will serve as a valuable resource for navi-

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

gating this complex topic and facilitate the improved use of the wealth of existing data to analyze and understand heterogeneous catalysis.

2 Overview of computational thermodynamics

This review aims to be self-contained and thus we begin with a brief overview of the relevant thermochemical properties and relationships. When working with thermochemical data related to heterogeneous catalysis or any other system, it is imperative to recognize that any thermochemically meaningful quantity will be a relative quantity that satisfies a mass balance. Here, we focus on chemical processes occurring at a fixed temperature and pressure; hence, we use Gibbs free energy ΔG as the relevant thermochemical potential. Thermodynamic equilibrium governs the maximum achievable conversion and selectivity for a chemical conversion process. The equilibrium constant K for a reaction is calculated from the Gibbs free energy change $\Delta_r G$ at a temperature T

$$K(T) = \exp\left(\frac{-\Delta_r G(T)}{RT}\right). \tag{1}$$

$\Delta_r G$ is the sum of the products of the free energies of formation $\Delta_f G_i$ of species i and their stoichiometric coefficient ν_i

$$\Delta_r G = \sum_{i=1}^{N} \nu_i \Delta_f G_i(T) \tag{2}$$

Thus, it is crucial to accurately know the thermochemical quantities of species participating in the reactions to make quantitative predictions with kinetic models. In microkinetic models (gas-phase or surface reaction mechanism), thermophysical properties of all species and intermediates are required to determine the equilibrium constants of the elementary steps. Commonly, an elementary reaction is specified in the forward direction with rate constant k_{fwd} and the rate constant for the reverse direction k_{rev} is determined from the concentration equilibrium constant K_c,

$$k_{rev} = \frac{k_{fwd}}{K_c} \tag{3}$$

which ensures microscopic reversibility and thermodynamic consistency. The free reaction energy is related to the enthalpy $\Delta_r H$ and entropy $\Delta_r S$ of reaction:

$$\Delta_r G(T) = \Delta_r H(T) - T\Delta_r S(T) \tag{4}$$

The reaction enthalpy is calculated from the enthalpy of formation $\Delta_f H_i(T)$, also known as the heat of formation, of the species,

$$\Delta_r H(T) = \sum_{i=1}^{N} \nu_i \Delta_f H_i(T) \tag{5}$$

where $\Delta_f H_i(T)$ is a critical thermochemical quantity since it captures the majority of the influence of chemical bonds. The enthalpy of formation is a relative quantity, meaning it is always defined relative to other species or reference states, whereas entropy and heat capacity are properties of an individual species. Defining a set of reference species is necessary to obtain a set of internally consistent $\Delta_f H$ for all species. The thermodynamic community has defined the elements in their standard configuration as the standard reference species[85], so the enthalpy of formation of a species is defined as a formation reaction from the elements in their most stable form at a constant temperature. For example, for C_2H_6

$$2C_{(graphite)} + 3H_2 \xrightarrow{\Delta_f H} C_2H_6 \tag{6}$$

where the reference species are graphite and H_2. Accordingly, the $\Delta_f H$ of the standard reference species are 0 since it is a null reaction. Experimental $\Delta_f H$ can only be obtained directly for a limited set of species such as H_2O (oxidation of H_2) and CO_2 (oxidation of graphitic C). In nearly all other cases, $\Delta_f H$ are derived from other measured reaction enthalpies, using the fact that the enthalpy is a state variable. Hess's law states that $\Delta_f H$ of a species is invariant to the reaction through which it is formed. This law can be exploited since it is relatively easy to measure the combustion enthalpy of C_2H_6 in a bomb calorimeter.

$$C_2H_6 + 3.5O_2 \xrightarrow{\Delta_r H^{combust.}} 2CO_2 + 3H_2O \tag{7}$$

Alternatively, this reaction can be written as:

$$\underbrace{2CO_2 + 3H_2O - 3.5O_2}_{\text{references}} \longrightarrow \underbrace{C_2H_6}_{\text{target}} \tag{8}$$

where the reference species to form our target species are located on the reactant side. If $\Delta_f H$ of O_2, CO_2, and H_2O are known ($\Delta_f H_{O_2} = 0\,\mathrm{kJ\,mol^{-1}}$, $\Delta_f H_{CO_2}$ and $\Delta_f H_{H_2O}$ can directly be measured), it is simple to back out $\Delta_f H_{C_2H_6}$ with the measured $\Delta_r H^{combust.}$:

$$\underbrace{\Delta_f H_{C_2H_6}}_{\text{target}} = \Delta_r H^{combust.} + \underbrace{\sum_{i \neq P} \nu_i \Delta_f H_i}_{\text{references}} \tag{9}$$

where ν_i is the stoichiometric coefficient of species i (e.g., $\nu_{CO_2} = -2$, $\nu_{O_2} = 3.5$ in Eq. (8)). Fig. 1 shows the thermodynamic cycle to determine $\Delta_f H_{C_2H_6}$.

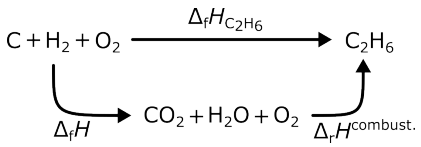

Fig. 1 Thermochemical cycle to determine the $\Delta_f H_{C_2H_6}$ from the combustion enthalpy.

However, $\Delta_f H$ of many species have never been measured and probably will never be measured, especially for highly reactive species such as radicals and other short-lived intermediates. Consequently, QM methods are used instead to compute $\Delta_f H$ of species that cannot be measured. In all equations, we use the general QM instead of DFT since all methods can be used with any level of theory of electronic structure data. QM methods can provide only electronic energies of individual species that are internally referenced to the QM reference frame of the software, which can then be used to compute the reaction enthalpies and finally deduce $\Delta_f H$. To determine $\Delta_f H$ of a target molecule P from

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

an electronic structure method, we form a hypothetical reaction from a set of reference species A, B, and C with known $\Delta_f H$.

$$a\mathrm{A} + b\mathrm{B} + c\mathrm{C} \longrightarrow \mathrm{P} \tag{10}$$

The first step is to compute the reaction enthalpy using the zero-point corrected QM energies (E_i).

$$\Delta_r H^{\mathrm{QM}} = \underbrace{E_\mathrm{P}}_{\text{target}} + \underbrace{\sum_{i\neq \mathrm{P}}^{N} \nu_i E_i}_{\text{references}} \tag{11}$$

With $\Delta_r H^{\mathrm{QM}}$, we can now back out $\Delta_f H$ of the target:

$$\underbrace{\Delta_f H_\mathrm{P}\,(0\mathrm{K})}_{\text{target}} = \Delta_r H^{\mathrm{QM}} - \underbrace{\sum_{i\neq \mathrm{P}}^{N} \nu_i \Delta_f H_i\,(0\mathrm{K})}_{\text{references}} \tag{12}$$

$$\underbrace{\Delta_f H_\mathrm{P}\,(0\mathrm{K})}_{\text{target}} = \underbrace{E_\mathrm{P}}_{\text{target}} + \underbrace{\sum_{i\neq P}^{N} \nu_i E_i}_{\text{references}} - \underbrace{\sum_{i\neq P}^{N} \nu_i \Delta_f H_i\,(0\mathrm{K})}_{\text{references}} \tag{13}$$

Eq. (12) is identical to Eq. (9); the only difference is the source of the reaction enthalpy. Since the reaction is hypothetical, it is possible to use any stoichiometrically balanced reaction. QM calculations are performed at a temperature of 0 K, so $\Delta_f H$ derived from the QM reaction enthalpy are at 0 K, too. They can be converted to $\Delta_f H$ at 298 K by applying temperature corrections[86]. *In principle, all the approaches described in this work are mathematically identical to Eq.* (13)*; they only differ in the choice of reference species and the value of $\Delta_f H$ of the reference species.*

2.1 Anchors, references, and thermochemical networks

Enthalpies are always derived from a set of reference species, so without a common set of references, the values of enthalpies are meaningless. When building microkinetic models, we are not only interested in the enthalpies or Gibbs free energies of a single species but rather the entire reaction pathway or mechanism. In heterogeneous catalysis, this includes the adsorption of gas-phase reactants, a series of surface reaction steps, and the desorption of gas-phase products. Thus, it is necessary to align the enthalpies of all species to the same set of reference species, which creates a *thermochemical network*. In this work, we introduce the term *anchor species* as a special case of these reference species that form the basis of a thermochemical network. The introduction of this term helps to resolve ambiguity in approaches where multiple or non-standard references are used. As we will show, the anchor species of a thermochemical network will have a relative enthalpy of zero within their own reference frame. In contrast, reference species may have zero or non-zero values.

The gas-phase thermodynamic community has defined the enthalpies of formation $\Delta_f H$ (see Eq. (6)), which use the elements in their most stable configuration as the reference thermochemical species[85]. By referencing all species to the standard set of anchor species (*i.e.* the elements in their IUPAC standard states[85]), it is possible to integrate data from multiple sources, e.g., experimental and theoretical determinations with various degrees of accuracy, into a *global* thermochemical network. There are many highly accurate QM methods available for gas-phase molecules (e.g., W4[87], HEAT[88], focal point[89], and ANL0[90]) that are capable of computing reaction enthalpies with an accuracy of $\pm 1\,\mathrm{kJ\,mol^{-1}}$. With these accurate reaction enthalpies, the gas-phase chemistry community has established standard methods to convert the electronic energies from QM into highly accurate $\Delta_f H$ in thermochemical networks, which can be created in two ways: *sequential* or using the *Active Thermochemical Tables approach*[76,91]. We briefly describe these two different approaches in this review, but the reader is referred to refs 86,91,92 for a complete description.

In the sequential approach, a *single* species is added by referencing it to species already in the thermochemical network. These species can be the elements in their standard states (anchors) or species that are connected to the elements, *i.e.* their $\Delta_f H$ is known (references). The expansion of the thermochemical network is most commonly done one species at a time, which introduces a few issues. Foremost, it requires the constructor of the sequential compilation to choose at the onset which species to add and in which order. Imagine that there are two species with similar quality thermochemical determinations (e.g. two different reaction enthalpies forming the target from references) that relate both to the existing compilation. The constructor of a sequential compilation must arbitrarily choose which to add first, artificially diminishing the impact that the second has on all subsequent determinations. Once a species is added, values for its enthalpy and partition function are adopted, and all additional species that include a reference to that species, even indirectly, depend explicitly upon the previous values. The choice of partition function determines how a species thermochemical information is converted to the reference temperature of 298.15 K under standard conditions at which the thermochemical network is solved. Thus, partition functions can limit the accuracy as many determinations occur at temperatures far from the reference temperature. The third issue is bias because constructors have to weigh all the available information when choosing what values to recommend for a given reaction enthalpy. Frequently, the constructors adopts a single "best" determination to add a new species to the network, discarding the remaining information. Importantly, new information that becomes available after a value has been adopted *cannot* be included without redoing the entire compilation of the network, which is the largest problem with the *sequential* compilation. By the time a sequential network is compiled there is likely better or contradictory information that cannot be included without introducing inconsistencies. Examples of these sequential thermochemical networks are the JANAF tables[74] or the NIST Chemistry WebBook[73].

The dynamic ATcT[75,76,92] approach overcomes the pitfalls of the adopt-and-freeze sequential approach. At the time of compilation of the thermochemical network, a set of partition functions is chosen for each species. The *entire* thermochemical network is constructed by fitting trial $\Delta_f H$ to that network using a weighted least squares regression[92]. All determinations within the net-

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

work are then checked against the trial solution using the "worst offenders" algorithm. The determination that is least consistent with the trial $\Delta_f H$ has their uncertainty increased marginally. This procedure is repeated until the entire thermochemical network consists of $\Delta_f H$ with uncertainties that are internally consistent. The ATcT approach uses all data available at the time of solution and can be painlessly updated with new information, including additional *ab initio* or experimental data, new species, or improved partition functions for species. Thus, the ATcT incorporates all available data from experimental and theoretical into a single thermochemical network. This approach can determine $\Delta_f H$ with an astonishing sub kJ mol^{-1} accuracy. The current version of the ATcT contains 3000 gas-phase species[75], and it can easily be accessed at `https://atct.anl.gov/`. Our current study uses the ATcT database as the global thermochemical network. Accordingly, we will use the terms ATcT or global thermochemical network interchangeably in the remainder of this paper.

In computational catalysis it is more common to create a self-consistent *local* thermochemical network. A local thermochemical network may be anchored to any valid set of anchor species. These anchor species can be part of the reaction mechanism or additional species not present in the mechanism, and they can even be abstracted into elemental chemical potentials. In the context of *ab initio* thermodynamics, we can think of these reference species as being in a reservoir that is in equilibrium with our target species (adsorbate, gas-phase, or surface)[93]. The reactions to form the target from the reference species can take or donate these references to the reservoir as an isothermal and isobaric process[94]. This work details the numerous approaches to construct these local networks and the approaches that integrate the formation enthalpies of adsorbates into the *global* thermochemical network by anchoring the species to the standard reference set according to the IUPAC definition.

2.2 Thermochemistry in heterogeneous catalysis

The gas-phase thermochemical data from the global thermochemical network is relevant for catalysis because it determines the overall thermochemical equilibrium of a catalyzed gas-phase reaction. However, no $\Delta_f H$ of adsorbates are included in the ATcT, except *H on Pt(111). The lack of adsorbates in these databases is due to the structural complexity of the active site motifs, finite coverage effects, the accuracy of electronic structure calculations for adsorbates, and the limited availability of accurate experimental values. The $\Delta_f H$ of the adsorbate depends on the active site (e.g., metal (oxide) facet), which drastically increases the space of possible species to integrate into a thermochemical network. Additionally, the atomic-scale structure of surfaces may vary even for a single facet, resulting in multiple possible binding sites with distinct $\Delta_f H$.

The accuracy of electronic structure calculations for gas/solid catalysis is typically limited to DFT because of the prohibitive computational cost of more advanced methods. Wellendorff et al.[67] benchmarked predicted reaction energies for the adsorption of gas-phase molecules on transition metals with a range of different xc functionals with known experimental values[95] and obtained the best results for the BEEF-vdW functional with uncertainties of ± 30 kJ mol^{-1} (mean absolute error compared to the benchmark dataset)[67]. Some studies employed higher levels of theory such as combinations of GGA and meta-GGA functionals[96,97], embedding techniques[98–100], random-phase approximation (RPA)[48,101–105], composite DFT/MP2/CCSD(T) methods[106–109], CCSD(T)[47], and Quantum Monte Carlo[46,110] to derive more accurate adsorption enthalpies. A review of beyond DFT methods is provided by Sauer[111]. However, even with high-level methods, heterogeneous systems typically require pseudopotentials and other numerical approximations, leading to lower accuracy than what can be achieved for gas-phase systems. Another source of parametric uncertainty is the partition function used to compute the entropy and heat capacity, which are often approximated by assuming an ideal gas, harmonic oscillator, or hindered translator/rotor[112–114]. The errors are typically smaller than errors in enthalpy but can still be substantial, as shown by more advanced techniques that account for anharmonicity[115–117].

Experimental adsorption and reaction enthalpies can be significantly more accurate than DFT data. Adsorption enthalpies can be derived from temperature-programmed desorption (TPD) spectra or measured using single-crystal adsorption calorimetry (SCAC). An excellent review on SCAC is provided by Campbell[118]. Determining $\Delta_f H$ of the adsorbates anchored to the global thermochemical network from experimental adsorption enthalpies is straightforward. For associative adsorption, the adsorbate is directly referenced to the gas-phase precursor; for example, the adsorption enthalpy of CH_4 on Pt(111) was measured by Tait et al.[119] via TPD:

$$CH_4 + {}^* \longrightarrow CH_4{}^* \qquad \Delta_r H(63\,\mathrm{K}) = -15\,\mathrm{kJ\,mol^{-1}}$$

Assuming that $\Delta_f H_{Pt(111)}$ is zero (a common assumption that we will discuss later), $\Delta_f H_{CH_4{}^*}$ can be calculated directly with $\Delta_f H_{CH_4}$ from the ATcT database in the thermochemical cycle in Fig. 2. Temperature corrections have to be applied to convert between the experimental and the reference temperature. However, many species dissociate easily once adsorbed or do not have a stable gas-phase precursor, requiring specialized precursors to measure their adsorption energy. For example, Karp and coworkers[120,121] used CH_3I as a precursor to determine $\Delta_f H_{^*CH_3}$

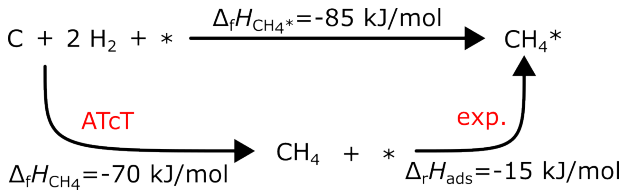

Fig. 2 Thermodynamic cycle to compute the formation enthalpy of $CH_4{}^*$ on Pt(111) from the experimental enthalpy of adsorption at 63 K.

$$CH_3I + 2\,{}^* \longrightarrow {}^*CH_3 + {}^*I \quad \Delta_r H(320\,\mathrm{K}) = -212\,\mathrm{kJ\,mol^{-1}}$$

and measured a reaction enthalpy for the dissociative adsorption on Pt(111) of -212 kJ mol^{-1} with SCAC[120]. $\Delta_f H_{^*CH_3}$ can then

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

be derived using a thermodynamic cycle as displayed in Fig. 3. The experimental determination of the accurate $\Delta_f H$ of adsor-

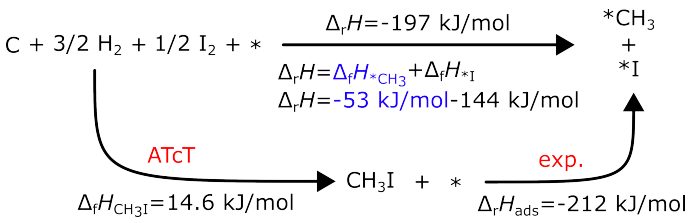

Fig. 3 Thermodynamic cycle to determine the $\Delta_f H*_{CH_3}$ on Pt(111) adapted from ref. 121. $\Delta_f H*_I$ ($-144\,\mathrm{kJ\,mol^{-1}}$) was obtained from a different experiment. The figure is adapted from Karp et al.[121].

bates is challenging since single-crystal adsorption calorimeters are expensive and intricate systems. Creating single crystals and ensuring no transformation during the operation is challenging. Additionally, $\Delta_f H$ of the co-adsorbates have to be known with high accuracy[122], and coverage effects caused by self- or cross-interaction can distort the measurement[120,121]. Another challenge for more complex adsorbates is finding suitable gas-phase precursors and preventing unwanted dissociation of the adsorbate. Due to all these factors, only a few dozen experimental adsorption enthalpies exist for a limited number of systems, summarized by Silbaugh and Campbell[95]. Consequently, it is fair to assume that most adsorbate $\Delta_f H$ will never be experimentally determined, and they have to be derived from QM data instead.

3 Definition of terms and variables

All pertinent definitions and associated notation are collected and concisely summarized in this section. We follow existing definitions and notations from the gas-phase chemistry literature where possible. Throughout this work, we assume that all thermochemical quantities are evaluated at 0 K, consistent with computational chemistry methods. For simplicity and readability, we drop the temperature in the equations below and only indicate if a quantity is evaluated at a different temperature. However, the definitions hold at all temperatures, and all mathematical relationships are valid as long as all quantities are evaluated at a consistent temperature. We also note that the existing computational catalysis literature may use different terms and contains many ambiguities. Thus, the definitions and notation below are inconsistent with some prior studies in computational catalysis, but are largely consistent with the gas-phase literature. Table 1 summarizes all the quantities used in this review. We provide a notation for a single species whenever possible and a linear algebra notation for a set of species (*e.g.*, full mechanism). The general notation uses a Δ for all relative quantities, with the subscript on the Δ indicating whether it is a formation f or reaction enthalpy r. The subscript on the H indicates the identity of the species, while the superscript on the H denotes the method used. We introduce another subscript A on the Δ that indicates whether the enthalpy is relative to a non-standard set of anchor species. For example "$\Delta_A H_{*CO}$ where $A = [CH_4,\ H_2O,\ H_2,\ Pt(111)]$" would fully specify an enthalpy of *CO adsorbed on Pt(111) anchored to the DFT energies of CH_4, H_2O, and H_2 and Pt(111). By definition, $\Delta_A H$ of the anchor species A are zero in this local thermochemical network. The specification of the subscript on the Δ is critical when reporting enthalpies of adsorbed species relative to non-standard anchors since, as shown later, the absolute values of the enthalpies vary widely with different anchor species. In this work, we reserve the name of formation enthalpies to its agreed upon definition by the chemistry community. Only enthalpies that are relative to the standard IUPAC anchors, either directly or indirectly, and, thus, are appended to the global thermochemical network are called enthalpies of formation.

4 Types of tasks in thermochemistry

The case study clarifies that diverse and often inconsistent datasets are prevalent in heterogeneous catalysis. Here, we seek to classify the various tasks that are often performed to make it more convenient to work with these quantities and correct or combine the values consistently. All tasks aim to construct a consistent thermochemical network that can be used to compute reaction energies that will be used as inputs to microkinetic models or other analyses. We broadly classify these into tasks of "comparison and conversion", which involve only arbitrary modifications of data such that **all relative quantities within a dataset remain unchanged**, and approaches for "combining and correcting" energies, which involve **changes to relative quantities within a dataset**. The first class of tasks aims to create a *local* thermochemical network with anchors that do not correspond to the standard states, whereas the latter class of tasks integrates the data into a *global* thermochemical network with the standard anchors that facilitates combining data from multiple sources (gas-phase species, experimental data, etc.). The tasks for combining and correcting require decisions and compromises for selecting or combining inconsistent data. Different decisions and compromises may yield different reaction energies. It is further possible to divide the approaches into two categories. Most described methods use an **atomic reference basis set**, where a single reference/anchor species is selected for every element. The basis set terminology was first introduced by Mhadeshwar et al.[123] for computational catalysis. Another approach has recently emerged that uses a **molecular fragment reference basis set**, where a reference species is defined for a particular feature of the target, *e.g.*, bond type.

This work aims to provide a standardized way of considering the various compromises involved with each approach and provide researchers with convenient mathematical tools for implementing them. Fig. 4 summarizes the different approaches graphically. All methods are discussed assuming $C_xH_yO_z$ molecules and adsorbates, but it is straightforward to extend the methods to an arbitrary number of chemical elements (*e.g.*, N or S) by adding additional reference species. Examples of all methods are provided in tutorial Jupyter Notebooks included in the supplementary information and in ref 124.

4.1 Case Study

We choose the oxidative dehydrogenation of ethane on Pt(111) using the mild oxidant CO_2 as a case study to test the different ap-

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Symbol	Definition
E_i	Zero-point corrected electronic (DFT) energy of a species i
$\Delta_\mathrm{f}H_i(T)$	Enthalpy of formation of a species i at temperature T anchored to the global thermochemical network (here ATcT)
$\Delta_\mathrm{A}H_i(T)$	Enthalpy of a species i at temperature T relative to the anchor species A in a local thermochemical network
$\Delta_\mathrm{r}H_{i,ads}(T)$	Enthalpy of adsorption (or enthalpy of the adsorption reaction) of species i at temperature T
$\Delta_\mathrm{r}H(T)$	Reaction enthalpy of a stoichiometrically balanced reaction (gas-phase, surface, and adsorption reaction) calculated from enthalpies of formation. This quantity is free of any anchor/reference species.
$\Delta_\mathrm{r}H^\mathrm{QM}$	Enthalpy of reaction j calculated only from DFT energies ($\Delta_rH^\mathrm{QM} = \sum \nu_i E_i$)
$\underline{\nu}$	Stoichiometry vectors with stoichiometry coefficients ν_i of each species i in a reaction
$\underline{\underline{\mathbf{M}}}$	Stoichiometry matrix that encompasses a set of stoichiometry vectors (species × references)
$\underline{\underline{\mathbf{N}}}$	Elemental composition matrix of a set of species (species × elements)
$\underline{\underline{\mathbf{F}}}$	Fragment composition matrix of a set of species (species × fragments, *e.g.*, bond types)
$\underline{\mathbf{H}}_\mathrm{f}(T)$	Vector of enthalpies of formation at temperature T
$\underline{\mathbf{H}}_\mathrm{A}(T)$	Vector of enthalpies relative to the defined anchor species A at temperature T
$\underline{\mathbf{H}}_\mathrm{f}^\mathrm{R}(T)$	Vector of enthalpies of formation of the reference species R at temperature T
$\underline{\mathbf{H}}_\mathrm{r}(T)$	Vector of reaction enthalpies at temperature T
$\underline{\mathbf{E}}$	Vector of zero-point corrected DFT energies
$\underline{\mathbf{E}}^\mathrm{A}$ or $\underline{\mathbf{E}}^\mathrm{R}$	Vector of zero-point corrected DFT energies of the anchor species A or reference species
Anchor	Anchor species are a special case of reference species because they form the foundation of a thermochemical network. The IUPAC standard states of each element are the anchor species for ATcT, NIST, and other global thermochemical networks, but anchor species can be arbitrarily defined for any local thermochemical network. The enthalpy of formation of an anchor species is always 0 because referencing it to itself creates a null reaction.
Reference	A species that is used to determine the reference chemical potential. We assume that the enthalpy of formation of a reference species is known, *e.g*, from the ATcT (global thermochemical network). The enthalpy of formation of the reference species is relative to the energy of an anchor species or other reference species.

Table 1 Definition of terms and variables. We drop the T for simplicity in the manuscript and only mention it if the temperature is relevant.

proaches to compute the enthalpies of formation of the species in a reaction mechanism and to construct the thermochemical networks.

$$C_2H_6 + CO_2 \rightleftharpoons C_2H_4 + CO + H_2O \; \Delta_rH(0\,\mathrm{K}) = 169.7\,\mathrm{kJ\,mol^{-1}}$$

This reaction is of interest to convert the large amounts of C_2H_6 in shale gas to C_2H_4 while at the same time reducing the greenhouse gas CO_2 to valuable CO. Pt and Pt-based alloys are promising catalysts for this reaction[125,126]. The considered mechanism is based on the pathways discussed by Jalid et al.[127], and we only consider one possible path of the complex reaction mechanism.

$$C_2H_6 + {}^* \rightleftharpoons C_2H_6{}^*$$

$$CO_2 + {}^* \rightleftharpoons CO_2{}^*$$

$$C_2H_6{}^* + {}^* \rightleftharpoons {}^*CH_2CH_3 + {}^*H$$

$${}^*CH_2CH_3 + 2^* \rightleftharpoons {}^*CH_2{}^*CH_2 + {}^*H$$

$$CO_2{}^* + {}^* \rightleftharpoons {}^*CO + {}^*O$$

$${}^*H + {}^*O \rightleftharpoons {}^*OH + {}^*$$

$${}^*OH + {}^*H \rightleftharpoons H_2O^* + {}^*$$

$$H_2O^* \rightleftharpoons H_2O + {}^*$$

$${}^*CO \rightleftharpoons CO + {}^*$$

$${}^*CH_2{}^*CH_2 \rightleftharpoons C_2H_4 + 2^*$$

We do not claim to provide new insights into this intricate reaction system, nor do we assume that the presented mechanism is complete. The case study serves purely as a demonstration purpose. Pt(111) was chosen as a model surface as it has a large amount of experimental data for $\Delta_\mathrm{f}H$ of adsorbates[95]. The $*$ indicates that the species is bound to the surface. If it is on the left-hand side of an atom, it indicates that the species is bound to the surface through this atom (*e.g.*, *OH). If it is on the right-hand side, it indicates that the species is physisorbed (e.g., H_2O^*). The DFT data used in this study is published in ref. 128. DFT calculations were performed with Quantum Espresso using the BEEF-vdW xc functional[129] on a (3×3) slab, which corresponds to a coverage of 1/9th monolayer. The reader is referred to ref. 128 for further details. Additionally, the RPBE[130] xc functional was employed to provide context on how to compare different functionals. Energetic data for the relevant species is summarized in Table S1.

Some methods construct the thermochemical network using only the species that are part of this mechanism and derive a set of internally consistent $\Delta_\mathrm{f}H$. However, in other methods, it is necessary to introduce a set of additional gas-phase or adsorbed species to determine $\Delta_\mathrm{f}H$ of the target species using hypothetical reactions and not the elementary reactions in the mechanism. These approaches connect the target to the existing global thermochemical network. In Table 2, we list all gas-phase species (either part of the mechanism or in a hypothetical reaction) and their $\Delta_\mathrm{f}H$ taken from the ATcT database (version 1.130[75,76]). We also assume that Pt(111) represents bulk Pt, which is why $\Delta_\mathrm{f}H_{\mathrm{Pt(111)}}$ is always $0\,\mathrm{kJ\,mol^{-1}}$. Technically, assuming that Pt(111) is bulk Pt is not in agreement with the IUPAC standard states since there is an

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

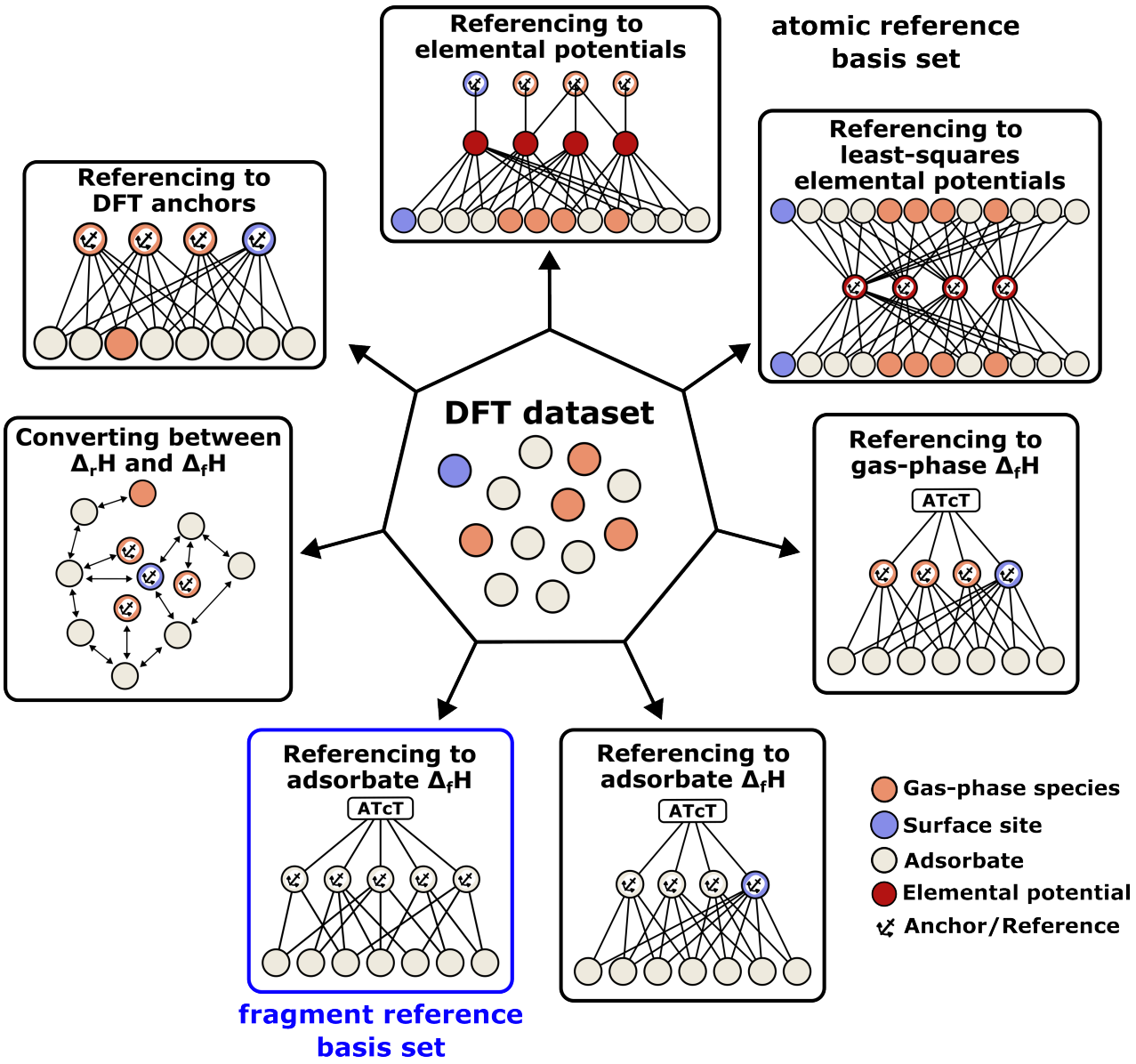

Fig. 4 Overview of the different methods to convert a DFT dataset to enthalpies of formation and construct a global or local thermochemical network. The methods can be grouped into atomic reference basis sets that conserve surface reaction energies from DFT and molecular fragment reference basis sets that exploit error cancellation. The methods can use only DFT energies, apply direct corrections to DFT energies, and consider experimental gas-phase species or adsorbate data.

enthalpy of formation to form Pt(111) from bulk Pt (proportional to the surface energy of Pt(111)). However, assuming the active site as a reference is the standard procedure in determining adsorption enthalpies from experiments or theory[95]. We note that in cases where multiple active sites are present, additional consideration is required to determine whether the difference in energies between active sites should be accounted for or not. Having multiple active sites with $\Delta_f H$ of zero is not thermodynamically consistent but may be appropriate if the relative concentrations of active sites are known or if surfaces are not assumed to be in equilibrium.

species	$\Delta_f H(0K)$ (kJ mol^{-1})	uncertainty (kJ mol^{-1})	source
*	0	exact	Assumption
H_2	0	exact	ATcT
O_2	0	exact	ATcT
CH_4	-66.549	± 0.044	ATcT
C_2H_6	-68.38	± 0.12	ATcT
H_2O	-238.902	± 0.022	ATcT
CO	-113.8	±0.026	ATcT
CO_2	-393.11	± 0.015	ATcT
C_2H_4	60.89	± 0.11	ATcT

Table 2 Enthalpies of formation of the gas-phase species at 0 K from the Active Thermochemical Tables (ATcT), version 1.130[75,92]. The vacant site * is Pt(111) in our case study.

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

4.2 Tasks for comparison and conversion of datasets

DFT total energies are anchored to a consistent zero energy, typically related to the vacuum level energy. However, the specifics will depend on the DFT code used, the xc functional, the pseudopotentials, and numerical settings. Nonetheless, the enthalpy $H(T,p)$ of a species (gas-phase or adsorbed) can be defined directly from a total DFT energy (E_{DFT}) by including the zero-point energy (E_{ZPE}) and a temperature correction $\delta H_i^{\mathrm{finite}}(T,p)$, which is based on the partition functions of the species.

$$H_i(T,p) = E_{\mathrm{DFT,i}} + E_{\mathrm{ZPE,i}} + \delta H_i^{\mathrm{finite}}(T,p) \tag{14}$$

$$= E_i + \underbrace{\delta H_i^{\mathrm{finite}}(T,p)}_{=0 \text{ at } 0\,\mathrm{K}} \tag{15}$$

For simplicity, we assume throughout this work that DFT energies E_i have already been zero-point corrected, and we assume a temperature of 0 K so that the finite-temperature correction can be neglected. The quantity $H_i(T,p)$ does not have a Δ as it is not an enthalpy of formation defined in a formation reaction; instead, it is defined based on the relative energies of the (valence) *electrons*. It is often more common in the catalysis literature to work with the Gibbs free energy $G(T,p)$, which is similarly defined as:

$$G_i(T,p) = E_i + \underbrace{\delta G_i^{\mathrm{finite}}(T,p)}_{=0 \text{ at } 0\,\mathrm{K}} \tag{16}$$

We reiterate that in this work, we are only evaluating the state variables at 0 K since the focus of this review is solely on species-interdependent enthalpic contributions to the Gibbs free energy of formation. At 0 K,

$$G_i(0\,\mathrm{K}) = H_i(0\,\mathrm{K}) = E_i. \tag{17}$$

All DFT energies from a consistent set of calculations create a local thermochemical network anchored to the same QM zero of energy. This local thermochemical network can be used to determine the relative quantities of interest for the researcher to evaluate the energetics of possible pathways of a mechanism [131–133]. However, using the DFT reference energy frame is inconvenient when comparing data from different sources, codes, and different levels of fidelity because the internal reference may change by tens or hundreds of eV depending on the code, pseudopotentials, or numerical settings selected. Therefore, converting the QM energies from the original internal electronic reference frame to an atomic reference basis set allows for easier data evaluation and comparison.

4.2.1 Referencing QM energies to elemental chemical potentials

A straightforward method to create a local thermochemical network of consistent $\Delta_{\mathrm{A}}H$ from the computed DFT energies is provided by the framework of *ab initio* thermodynamics [93,94]. In this framework, we formulate a reaction in which the target species is formed from the constituent elements to calculate $\Delta_{\mathrm{A}}H$ or the relative Gibbs free energy,

$$n_{\mathrm{C}}\mathrm{C} + n_{\mathrm{H}}\mathrm{H} + n_{\mathrm{O}}\mathrm{O} + n_{*} * \xrightarrow{\Delta_{\mathrm{A}}H} \mathrm{P}$$

where P is a generic $C_xH_yO_z$ gas-phase species or a $C_xH_yO_z{}^*$ adsorbate. We will start the derivations with the relative Gibbs free energy, which is more common in the *ab initio* thermodynamics literature. The Gibbs free energy G_i of a species is defined as:

$$G_i(T,p) = G_i(0\,\mathrm{K},p) + \delta G_i^{\mathrm{finite}}(T,p) \tag{18}$$

$$G_i(T,p) = E_i + \delta G_i^{\mathrm{finite}}(T,p) \tag{19}$$

When defining our atomic reference basis set A as the DFT energies of the elements in the reaction above, we can simply determine the Gibbs free energy $\Delta_{\mathrm{A}}G$ in this reference frame from the reaction free energy.

$$\Delta_{\mathrm{A}}G(T,p) = \Delta_{\mathrm{r}}G(T,p) \tag{20}$$

$$\underbrace{\Delta_{\mathrm{A}}G(T,p)}_{\text{target}} = \underbrace{G_{\mathrm{P}}(T,p)}_{\text{target}} - \underbrace{\sum_k^{N_{\text{elements}}} n_k G_k(T,p)}_{\text{references}} \tag{21}$$

The Gibbs free energy of the elements G_k is equal to the elemental chemical potential $\mu_k(T,p)$, which is more commonly used in the literature.

$$\underbrace{\Delta_{\mathrm{A}}G(T,p)}_{\text{target}} = \underbrace{G_{\mathrm{P}}(T,p)}_{\text{target}} - \underbrace{\sum_k^{N_{\text{elements}}} n_k \mu_k^{\mathrm{A}}(T,p)}_{\text{references}} \tag{22}$$

n_k is the number of atoms of the element k in species i (identical to the stoichiometric coefficient in the reaction above). Since we focus only on a temperature of 0 K, we can use Eq. (17) to reduce Eq. (21) to

$$\Delta_{\mathrm{A}}G(0\,\mathrm{K}) = \Delta_{\mathrm{A}}H = E_{\mathrm{P}} - \sum_k^{N_{\text{elements}}} n_k \mu_k^{\mathrm{A}}. \tag{23}$$

In this *ab initio* thermodynamics framework, $\Delta_{\mathrm{A}}H$ of the target is anchored to the chemical potentials of the elements μ_k, which we need to determine. Theoretically, it is possible to use the DFT energies of the elements as the chemical potentials. However, it is impractical to reliably determine the chemical potentials (or electronic energies) of the standard states of the elements with DFT. For example, the graphite formation energy is sensitive to vdW forces, which are not accurately accounted for in many functionals, and the O_2 energy requires an accurate treatment of the triplet spin state of oxygen, which is difficult at the DFT level of theory and depends on the spin state. Total atomization energies offer a convenient alternative in principle, but in practice, they suffer from similar issues. Treatment of bare atoms in DFT is generally inaccurate due to unpaired electrons. Also, it requires careful convergence of spin states that can cause ambiguity in the results due to potentially large numerical errors. Calculating accurate atomization energies demands the highest level of theory and presents the most demanding calculations. To avoid these challenges, we can derive the chemical potentials of the elements via two different approaches: (i) by hand-picking closed-shell reference species that are supposedly more accurate and use them to

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

calculate the chemical potentials in a sequential approach, or (ii) by formulating a linear algebra problem that can be solved via linear regression, avoiding explicitly defined reference species. The first approach is described in this section, and the second is described in Section 4.2.3.

In principle, the choice of reference species is arbitrary, but there are some constraints. Reference species have to be chosen so that there is a species for every element. For example, we can use A=[CH_4, H_2O, H_2] as closed-shell anchor species to determine the chemical potentials. It is assumed that these reference species are available in thermodynamic reservoirs[93]. These reservoirs are connected with the unit cell of the target molecule/adsorbate/structure, and the reference species can be drawn from the reservoirs or vanish in them without affecting the pressure or temperature of the target[94]. The chemical potentials are then derived, assuming a thermoneutral reaction to form the element from the closed-shell reference species.

$$CH_4 - 2H_2 \longrightarrow C$$

The chemical potential μ_C^A is then calculated from the DFT energies and other μ_k^A are derived similarly.

$$\mu_C^A = E_{CH_4} - 2E_{H_2} \tag{24}$$

$$\mu_O^A = E_{H_2O} - E_{H_2} \tag{25}$$

$$\mu_H^A = 0.5E_{H_2} \tag{26}$$

We have now abstracted the DFT energies of the chosen anchor species into the elemental chemical potential. We can also assume that the Pt(111) slab is one entity and define the chemical potential of this reference as the DFT energy of the surface site $\mu_*^A = E_{Pt(111)}$. $\Delta_A H_P$ referenced to the anchors A = [CH_4, H_2O, H_2, Pt(111)] is then defined as:

$$\underbrace{\Delta_A H_P}_{\text{target}} = \underbrace{E_P}_{\text{target}} - \underbrace{\sum_k^{N_{\text{elements}}} n_k \mu_k^A}_{\text{references}} \tag{27}$$

It immediately follows that $\Delta_A H_k = 0$. The linear algebra notation of Eq. (27) is,

$$\underline{\mathbf{H}}_A = \underline{\mathbf{E}} - \underline{\underline{\mathbf{N}}}\,\underline{\mu}^A \tag{28}$$

where $\underline{\underline{\mathbf{N}}}$ is the elemental composition matrix (number of molecules × number of elements) of the target species.

$$\underline{\underline{\mathbf{N}}} = \begin{array}{c} \text{Elements} \\ \begin{array}{cccc} \text{C} & \text{H} & \text{O} & * \end{array} \\ \begin{bmatrix} 2 & 6 & 0 & 0 \\ 2 & 6 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \end{array} \begin{array}{l} C_2H_6 \\ C_2H_6{}^* \\ \vdots \end{array} \quad \text{Species} \tag{29}$$

For a set of reference species to determine the chemical potentials, it is also possible to determine the chemical potentials by using the inverse of the elemental composition matrix of the anchors.

$$\underline{\mu}^A = \underline{\underline{\mathbf{N}}}^{A^{-1}}\,\underline{\mathbf{E}}^A \tag{30}$$

For simplicity, we integrated the entire slab directly into the elemental composition matrix, which is convenient when determining only $\Delta_A H$ of adsorbate and gas-phase species. It is possible to parameterize a microkinetic model with the derived $\Delta_A H$ from this method[134]. However, the chemical potential of the slab can also be broken down into every atom, which is useful when investigating mixed transition metal, metal oxide, or metal carbide surfaces[94,135–137]. Explicitly defining the chemical potentials of all atoms is often used to derive surface free energies γ[94].

$$\gamma_P(T,p,...) = \frac{1}{2A}\left[G_P(T,p,...) - \sum_k^{N_{\text{elements}}} n_k \mu_k(T,p,...)\right] \tag{31}$$

The surface free energies can be used to construct phase diagrams and compare the stability of materials for catalytic and electrocatalytic systems under specific reaction conditions[135–142]. However, they may also be related to specific reaction conditions such as temperatures, environment, applied potentials, or pH. In the cases where the chemical potential is evaluated at different temperatures than 0 K, it is necessary to work with the Gibbs free energy of formation in Eq. (21) and consider finite temperature and other corrections. An example is the computational hydrogen electrode (CHE), which also uses the concept of elemental chemical potentials[143–145]. For electrocatalytic systems, an additional reference potential is required for the proton-electron pair, defined as the reversible hydrogen electrode (RHE).

$$H^+ + e^- \rightleftharpoons \frac{1}{2}H_2$$

The proton-electron pair chemical potential is then defined via

$$\mu_{H^+} + \mu_{e^-} = 0.5\mu_{H_2} - eU \tag{32}$$

$$\mu_{H^+} + \mu_{e^-} = 0.5\left(E_{H_2} + \delta G_{H_2}^{\text{finite}}\right) - eU \tag{33}$$

where U is the applied potential and e the elementary positive charge. The chemical potential of H_2, μ_{H_2}, is the sum of the electronic energy E_{H_2} and the finite free energy corrections $\delta G_{H_2}^{\text{finite}}$. With this definition of the chemical potential of the proton-electron pair, it is possible to evaluate, *e.g.*, the surface morphology or stability of alloys at different applied potentials.

It is important to note that the numerical values of the $\Delta_A H$ will vary drastically depending on the choice of anchor species used to derive the chemical potentials of the elements. Consequently, directly combining energetic data with different anchors is erroneous. To illustrate this variation, we chose three different sets of anchors corresponding to closed-shell molecules to determine the chemical potentials, which are $A = [CH_4, H_2O, H_2, Pt(111)]$, $B = [CO, H_2O, H_2, Pt(111)]$, and $C = [C_2H_6, CO, H_2, Pt(111)]$.

Fig. 5 shows the calculated enthalpy profiles using the different anchor species for the oxidative C_2H_6 dehydrogenation. Indeed, it can be seen that $\Delta_A H$ for the species are very different as seen by the offsets between the profiles on the y-axis. Thus, a naive comparison of energies from $\Delta_A H$ and $\Delta_B H$ would result

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

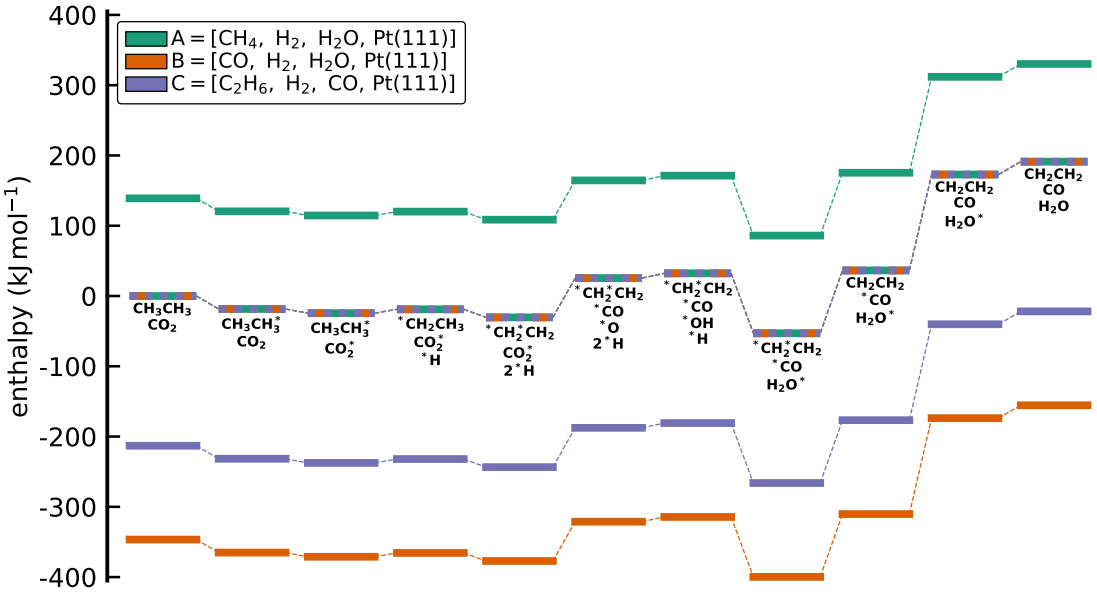

Fig. 5 Enthalpy diagram for the approach to reference the DFT energies to a set of anchor species through elemental chemical potentials. While the absolute values of the formation enthalpies depend on the set of anchor species, the reaction enthalpies for all elementary steps are conserved. This conservation can be seen when all lines are collapsed (multicolored line). No activation barriers are included in this diagram.

in a large discrepancy. However, the relative enthalpy differences between the adsorbates, *i.e.* the reaction enthalpies, which are ultimately the quantity that matters in microkinetic models, are identical. This is easily seen by setting the starting point in the enthalpy diagram to 0, which results in the collapse of the three different profiles. Thus, the reaction enthalpies predicted from DFT are always conserved, leading to identical predictions of a microkinetic model, regardless of the reference/anchor basis set.

For pedagogical purposes, we show explicitly that the reaction enthalpies are independent of the chosen references for any atomic reference basis set, *i.e.* when there is a single chemical potential for each chemical element. A relative reaction enthalpy can be calculated using $\Delta_A H$ derived from the chemical potentials, which is written as:

$$\Delta_r H = \sum_i^N \nu_i \Delta_A H_i \tag{34}$$

$$= \sum_i^{N_{products}} \nu_i \Delta_A H_i - \sum_i^{N_{reactants}} \nu_i \Delta_A H_i \tag{35}$$

$$= \sum_i^{N_{products}} \nu_i \left(E_i + \sum_{k \in products}^{K} n_k \mu_k^A \right) - \sum_i^{N_{reactants}} \nu_i \left(E_i + \sum_{k \in reactants}^{K} n_k \mu_k^A \right) \tag{36}$$

$$= \sum_i^{N_{products}} \nu_i E_i - \sum_i^{N_{reactants}} \nu_i E_i + \left(\sum_i^{N_{products}} \nu_i \sum_{k \in products} n_k \mu_k^A - \sum_i^{N_{reactants}} \nu_i \sum_{k \in reactants} n_k \mu_k^A \right) \tag{37}$$

$$= \sum_i^{N_{products}} \nu_i E_i - \sum_i^{N_{reactants}} \nu_i E_i \tag{38}$$

Therefore, for every stoichiometrically balanced reaction, it is evident that $k \in products$ and $k \in reactants$ will be equivalent and $\sum_k n_k \mu_k^A$ will cancel, making the choice of μ_k^A arbitrary. This invariance proves that as long as a dataset is anchored to the same set of atomic reference species, it is internally consistent and the relative quantities required for microkinetic models will be identical regardless of the anchors selected. The following techniques for comparison and conversion of enthalpies of formation exploit

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

this invariance, selecting "convenient" μ_k^{A} that allow for comparison of energies under different thermodynamic scenarios or between different levels of theory and for conversion of energies between relative and absolute reference states.

4.2.2 Referencing QM data to a set of anchor species

In the previous approach, we created a local thermochemical network where $\Delta_{\mathrm{A}}H$ are referenced to the elemental chemical potentials derived from a set of anchor species A. Here, we show that it is possible to cut the "middleman" of elemental potentials and use the anchor species directly to compute $\Delta_{\mathrm{A}}H$. As demonstrated above, the choice of reference species is in principle arbitrary, but there are some constraints that must be followed. One reference species must be chosen for every element to form an atomic basis set. As we will see in this section, the stoichiometry of anchor species must also not be linearly dependent. Common choices for the sets of reference species in computational catalysis are typically gas-phase molecules like CO, H_2, H_2O[146,147] or, CH_4,H_2, H_2O[148,149]. A vacant metal slab, *e.g.*, Pt(111), has to be added for adsorbed species. The construction of local thermochemical networks is frequently used in the literature, although anchor species often vary between studies and are not always explicitly stated. Referencing a single target species directly to a set of anchor species works as follows: We construct a hypothetical reaction to form the target (either gas-phase species or adsorbate) from the set of anchor species, *i.e.*, CH_4, H_2O, H_2, and Pt(111).

$$\nu_{\mathrm{CH_4}}\mathrm{CH_4} + \nu_{\mathrm{H_2O}}\mathrm{H_2O} + \nu_{\mathrm{H_2}}\mathrm{H_2} + \nu_{*}* \xrightarrow{\Delta_{\mathrm{A}}H} \mathrm{P}$$

When using the DFT values of the reference species as the anchor, it follows that $\Delta_{\mathrm{A}}H$ of the target referenced to this set of anchor species is equal to the reaction enthalpy of the reaction above.

$$\underbrace{\Delta_{\mathrm{A}}H_{\mathrm{P}}}_{\text{target}} = \Delta_{\mathrm{r}}H^{\mathrm{QM}} \tag{39}$$

We can compute this reaction enthalpy from the electronic energies.

$$\Delta_{\mathrm{r}}H^{\mathrm{QM}} = \underbrace{E_{\mathrm{P}}}_{\text{target}} + \underbrace{\sum_{i\neq\text{target}}^{N} \nu_i E_i}_{\text{references}} \tag{40}$$

To determine $\Delta_{\mathrm{A}}H_{\mathrm{P}}$, we can now substitute $\Delta_{\mathrm{r}}H$ in Eq. (39) with Eq. (40).

$$\Delta_{\mathrm{A}}H_{\mathrm{P}} = \Delta_{\mathrm{r}}H^{\mathrm{QM}} = E_{\mathrm{P}} + \sum_{i\neq\mathrm{P}}^{N} \nu_i E_i \tag{41}$$

The subscript A is the set of anchor species, in this case, $A = [CH_4, H_2O, H_2, \mathrm{Pt(111)}]$. $\Delta_{\mathrm{A}}H$ of our chosen anchor species are 0 because it is a null reaction to form the anchor species from themselves. Eq. (41) converts the zero-point corrected DFT energy of a single species into $\Delta_{\mathrm{A}}H$ referenced to the set of anchor species, A. This problem can also be cast as a linear algebra problem for a set of species, typically the entire reaction mechanism. The linear algebra notation using the definitions in Table 1 is as follows:

$$\underline{\mathbf{H}}_{\mathrm{A}} = \underline{\mathbf{E}} + \underline{\underline{\mathbf{M}}}\,\underline{\mathbf{E}}^{\mathrm{A}} \tag{42}$$

where $\underline{\mathbf{E}}$ is the vector of the zero-point corrected DFT energies, $\underline{\underline{\mathbf{M}}}$ is an $m \times n$ stoichiometry matrix with m anchor species and n reactions to form the target from the anchors (not to be confused with the elementary reactions in the mechanism), and $\underline{\mathbf{E}}^{\mathrm{A}}$ is the vector of the DFT energies of the anchor species. The stoichiometry matrix $\underline{\underline{\mathbf{M}}}$ can be constructed by hand, but it is also possible to use linear algebra to determine this from the elemental composition matrix $\underline{\underline{\mathbf{N}}}$. $\underline{\underline{\mathbf{N}}}$ an $m \times n$ matrix, but consists of the elemental composition of each species and is, thus, easier to construct and more general since it is independent of the anchor set. We have to construct an elemental matrix for the anchor species $\underline{\underline{\mathbf{N}}}^{\mathrm{A}}$ to determine $\underline{\underline{\mathbf{M}}}$.

$$\underline{\underline{\mathbf{N}}}^{\mathrm{A}} = \begin{array}{c} \text{Elements} \\ \begin{array}{cccc} \mathrm{C} & \mathrm{H} & \mathrm{O} & * \end{array} \\ \begin{bmatrix} 1 & 4 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ \mathrm{CH_4} \\ \mathrm{H_2} \\ \mathrm{H_2O} \\ \mathrm{Pt(111)} \end{array} \quad \text{Species} \tag{43}$$

To convert between the elemental composition matrix $\underline{\underline{\mathbf{N}}}$ of the species to the stoichiometry matrix $\underline{\underline{\mathbf{M}}}$ that forms the species from the anchor basis set, we can use the inverse of the elemental matrix of the reference species $\underline{\underline{\mathbf{N}}}^{\mathrm{A}^{-1}}$.

$$\underline{\underline{\mathbf{M}}} = -\underline{\underline{\mathbf{N}}}\,\underline{\underline{\mathbf{N}}}^{\mathrm{A}^{-1}} = \begin{array}{c} \text{References} \\ \begin{array}{cccc} \mathrm{CH_4} & \mathrm{H_2} & \mathrm{H_2O} & \mathrm{Pt(111)} \end{array} \\ \begin{bmatrix} -2 & 1 & 0 & 0 \\ -2 & 1 & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \end{array} \begin{array}{c} \\ \mathrm{C_2H_6} \\ \mathrm{C_2H_6}^{*} \\ \vdots \end{array} \quad \text{Species} \tag{44}$$

Note that this places an explicit constraint on any possible anchor set: The composition matrix $\underline{\underline{\mathbf{N}}}^{\mathrm{A}}$ must be invertible. This method of referencing DFT energies to a set of anchor species with their DFT values is commonly used. It is suitable for investigating a catalyst for a specific process[150–152], it can be used to screen across the material space[146,153–156], coverage effects on the thermophysical parameters can be included[147], it has been combined with uncertainty quantification[69,157], and it has also been applied to investigate electrocatalytic systems[158]. Since the method is identical to the elemental chemical potentials, it is also possible to directly combine it with the CHE model to calculate free energies at applied potentials[159]. This method is also the default for calculating $\Delta_{\mathrm{A}}H$ in the open-source CatMap[160] software package, although the notation and nomenclature used in the CatMap package and manuscript are different from the more consistent ones used in this work.

4.2.3 Referencing QM data to least-squares elemental chemical potentials

For a typical DFT dataset, there is a myriad of possible anchor species to calculate the chemical potentials. For example, we can calculate μ_O with different reference anchors, $A = [H_2O, H_2]$ and

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

$B = [\mathrm{CO}, \mathrm{CO_2}]$ through

$$\mu_{\mathrm{O}}^{\mathrm{A}} = E_{\mathrm{H_2O}} - E_{\mathrm{H_2}} = -55{,}799\,\mathrm{kJ\,mol^{-1}} \tag{45}$$

$$\mu_{\mathrm{O}}^{\mathrm{B}} = E_{\mathrm{CO_2}} - E_{\mathrm{CO}} = -55{,}878\,\mathrm{kJ\,mol^{-1}} \tag{46}$$

Clearly, different anchor species will lead to different values of the chemical potentials for the elements, which in turn affects the relative enthalpies as shown in Fig. 5. However, as demonstrated by Eq. (34), the reaction enthalpies of any stoichiometrically balanced reaction will be conserved in an atomic reference basis set. Instead of manually selecting the anchor species to determine the chemical potentials of the elements in a sequential approach, it is possible to exploit the invariance and select convenient numerical values for the chemical potentials. In particular, it is often convenient to select chemical potentials that minimize the (sum of squared) errors between relative enthalpies computed using different codes, levels of fidelity, or reference sets. As we will show, this can be achieved using linear algebra and least-squares regression and avoids the need to specify any explicit chemical species as anchors.

To calculate the least-squares anchor elemental potentials, we take the elemental composition matrix $\underline{\underline{\mathbf{N}}}$ that contains all the formation reactions of the target species from the constituent elements to derive the chemical potentials. Rather than selecting a set of explicit anchor species with defined stoichiometries, we seek the set of chemical potentials that minimizes the squared magnitude of the resulting relative enthalpies. The linear algebra formulation offers a convenient way to perform the minimization of squared relative enthalpies magnitudes, which is closely related to the least-squares regression problem and is given in Eq. (47).

$$\underline{\mu}^{\mathrm{LS}} = \left(\underline{\underline{\mathbf{N}}}^{\intercal}\underline{\underline{\mathbf{N}}}\right)^{-1}\underline{\underline{\mathbf{N}}}^{\intercal}\underline{\mathbf{E}} = \underline{\underline{\mathbf{N}}}^{+}\,\underline{\mathbf{E}} \tag{47}$$

The least-squares regression result can also be written as the product of the pseudo-inverse of the elemental composition matrix $\underline{\underline{\mathbf{N}}}^{+}$ and the energy vector $\underline{\mathbf{E}}$. Calculation of $\Delta_{\mathrm{LS}}H$ is then performed by replacing the chemical potentials in Eq. (47) with the expression in Eq. (28), leading to

$$\underline{\mathbf{H}}_{\mathrm{LS}} = \underline{\mathbf{E}} - \underline{\underline{\mathbf{N}}}\,\underline{\mu}^{\mathrm{LS}} = \underline{\mathbf{E}} - \underline{\underline{\mathbf{N}}}\,\underline{\underline{\mathbf{N}}}^{+}\,\underline{\mathbf{E}} \tag{48}$$

where we use the LS symbol to denote a "least-squares" anchor set that does not explicitly depend on specific molecular anchor species, but rather implicitly depends on the entire set of target species included (*i.e.* the species in the rows of $\underline{\underline{\mathbf{N}}}$).

The least-squares anchor set is particularly advantageous for comparing energies from different sources or approximations. When using specific molecular species as anchors A, any error associated with the species will propagate through $\Delta_{\mathrm{A}}H$ of the entire dataset. To illustrate this, Fig. 6a compares the relative enthalpies of species in the oxidative dehydrogenation of ethane for two different xc functionals, RPBE and BEEF-vdW. Relative enthalpies are computed using the elemental chemical potential approach with two different reference anchor sets and the least-squares anchor. Using the molecular anchors, the deviation between the two functionals varies widely depending on the choice of anchor species. Naive comparisons could lead to very different conclusions about the relative performance of these two xc functionals. On the other hand, using the entire reaction mechanism to determine the anchor chemical potentials by minimizing the sum of squared errors removes the arbitrary dependence on the choice of anchor species. When using these fitted chemical potentials, the scatter and, accordingly, the deviation between the xc functionals are significantly smaller. This behavior is also displayed in Fig. 6b, where $\Delta_{\mathrm{LS}}H$ derived from the RPBE and BEEF-vdW functional give similar results with only small deviations. Determining the chemical potentials via linear regression does not make the DFT energetics more accurate; all reaction energies will be unaffected, as shown in Eq. (34). However, it makes the energetic data from different xc functionals more comparable, as seen in Fig. 6b.

Using the least-squares chemical potentials as an anchor is similar to determining atomization energy corrections (AEC) often employed in the gas-phase community[161]. Known experimental atomization energies for species are used to determine the correction factors for the DFT energies. In a more generic way, this can be formulated as determining corrections to a set of energies to align them with known energies of higher fidelity. In the case of two DFT datasets obtained from two different functionals, we can either compare the relative enthalpies or determine atomic correction factors to align the energies of, *e.g.*, the RPBE functional to align with the BEEF energy values. In fact, it can be shown that the least-squares anchor is also the anchor set that minimizes the sum of squared deviation between two different datasets obtained from different sources. Thus, the minimum sum of squares chemical potentials can also be used directly to align data from different sources, *e.g.*, from different functionals such as BEEF-vdW and RPBE or different levels of theory.

$$\hat{\underline{\mathbf{E}}}^{\mathrm{RPBE}} = \underline{\mathbf{E}}^{\mathrm{RPBE}} + \underline{\underline{\mathbf{N}}}\,\underline{\underline{\mathbf{N}}}^{+}\left(\underline{\mathbf{E}}^{\mathrm{BEEF\text{-}vdW}} - \underline{\mathbf{E}}^{\mathrm{RPBE}}\right) \tag{49}$$

where $\hat{\underline{\mathbf{E}}}^{\mathrm{RPBE}}$ are the aligned energies of the RPBE functional to the BEEF-vdW functional. Eq. (49) allows to align the data from various sources to the same QM zero-of-energy, which is useful in data science and machine learning for providing an unbiased error between different levels of theory[97,162]. The method is mathematically similar to the Δ-ML approach, where the thermochemical quantity from on level of theory can be mapped onto another level of theory[163]. Here, we present how to do this using least-squares regression to determine elemental chemical potentials. However, the least-squares regression can be replaced with a machine learning model to regress other structural features using *e.g.* a LASSO (least absolute shrinkage and selection operator) optimization[164,165]. The downside of this approach is that the anchor chemical potentials will depend implicitly on all the species present in the dataset and technically need to be re-computed any time the network is expanded, similar to the ATcT[76].

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

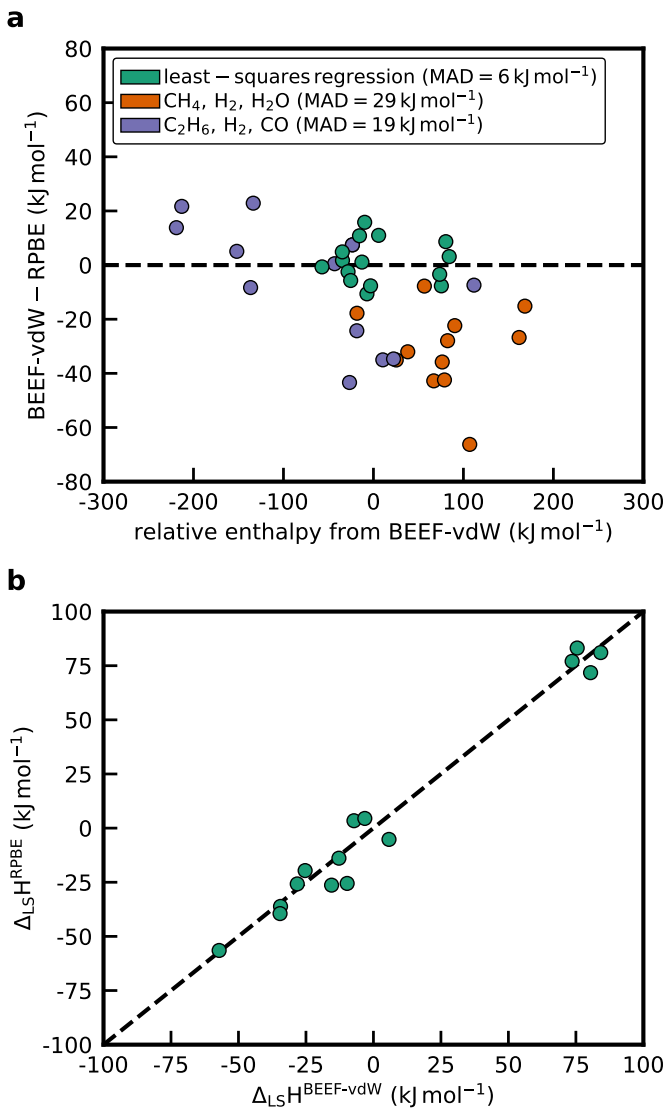

Fig. 6 a). Comparison of the difference between the enthalpies of formation from the BEEF-vdW and RPBE functional using different sets of references to determine the elemental chemical potentials: least-squares regression, $CH_4/H_2/H_2O$, and $C_2H_6/H_2/CO$. The mean absolute deviation (MAD) is reported in the legend. b) Aligning the two DFT datasets for the test case with different xc-functionals using linear regression to determine the elemental chemical potentials.

4.3 Approaches for correcting and combining datasets in atomic reference basis sets

The approaches presented above are convenient for converting energies into various frameworks such as *ab initio* thermodynamics and assist in comparing different energy sets while ensuring that all relative quantities for any reaction are conserved. Thus, using these approaches, all microkinetic models will be identical regardless of the choice of anchor species. However, the resulting reaction enthalpies may not be consistent with accurately known enthalpies of the overall gas-phase reaction from global thermochemical networks, such as the ATcT. For example, a reaction enthalpy of 191.2 kJ mol^{-1} is calculated for C_2H_6 dehydrogenation with CO_2 at 0 K using BEEF-vdW. This reaction enthalpy *is identical for all previously described approaches*. However, this value is in stark disagreement with the known reaction enthalpy of 169.7 kJ mol^{-1} determined from the accurate $\Delta_f H$ of the ATcT database. The difference results in an equilibrium constant K for the ODH reaction that is off by a factor of ∼20 for a typical reaction temperature of 873 K as shown in Fig. 7. Therefore, *the methods for comparison and conversion do not lead to microkinetic models that are thermodynamically consistent with the known $\Delta_f H$ of gas-phase species*. The methods described above are commonly used techniques for manipulating DFT energies in computational catalysis, and lay the foundation for subsequent approaches that enable datasets that are more consistent with known thermochemical information.

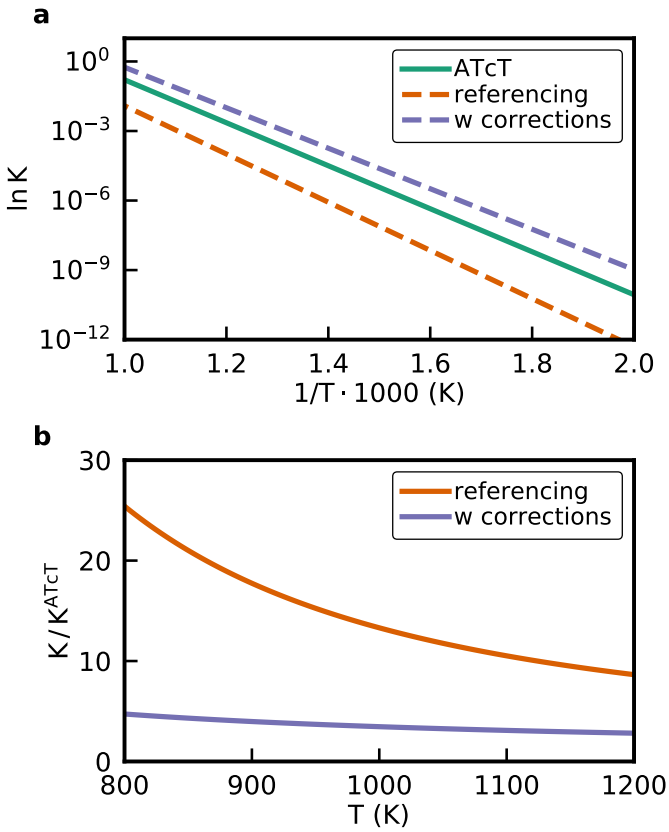

Fig. 7 a) Van't Hoff diagram of the equilibrium constant K derived from the enthalpies of reaction calculated from the ATcT database, the referencing approach, and the referencing approach with corrections. We assumed the same reaction entropy for all different cases. b) Ratio of the equilibrium constants from the computational chemistry methods compared to the exact equilibrium constant derived from the ATcT. Note that the inverse ratio of the referencing approach is shown.

The error in the gas-phase reaction energies is typically caused by the inaccuracy of common GGA xc functionals, which are on the order of 30 to 50 kJ mol^{-1} [67]. When computing the reaction enthalpy of surface reactions from DFT energies, it is often assumed that error cancellation improves the accuracy [69]. However, the degree of error cancellation can vary widely, depending on the specific reaction [128]. Thus, we hope for error cancellation for surface reaction energies, but it is not guaranteed that surface reaction energies are accurate [93]. Additionally, DFT calculations perform

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

poorly for gas-phase species due to the approximate treatment of electron exchange and correlation[166]. The most commonly used xc functionals in heterogeneous catalysis, such as BEEF-vdW and RPBE, are semi-empirical. They are specifically selected to predict experimental adsorption enthalpies, which reduces accuracy for gas-phase reaction energies[129].

Combining thermochemical data from various sources (typically of very different fidelity) is necessary to achieve better agreement with experimental gas-phase reaction enthalpies or more accurate $\Delta_f H$ of the adsorbates. This combination of data leads to changes in relative quantities in the thermochemical network that either implicitly or explicitly "correct" reaction enthalpies to be more consistent with higher-fidelity data. Schemes are available that apply these corrections to a few species, many species, or sometimes all species in the thermochemical network. Here, we classify approaches as "direct" correction schemes, where individual DFT energies are explicitly corrected, and "indirect" correction schemes, where corrections are made implicitly through error cancellation reactions. The direct methods tend to be simpler because they apply fewer corrections and can be understood intuitively, but they typically do not exactly reproduce all known quantities. In contrast, indirect methods are generally capable of reproducing known quantities exactly, but they do so at the expense of greater complexity and less transparency.

4.3.1 Corrections to the DFT energies of gas-phase species

It is possible to improve the accuracy of the overall reaction enthalpy by adjusting the DFT energies of individual gas-phase species. Correction factors include atomic energy corrections (AEC), bond additivity corrections (BAC), or corrections to fragments or specific molecules[161,167]. AECs and BACs are standard correction methods in the gas-phase community but are less frequently used in catalysis. In computational catalysis, correction factors have mainly been applied to specific molecules.

The most prominent introduction of molecular corrections in heterogeneous catalysis is the correction of the O_2 molecule, first introduced by Nørskov *et al.* to explain the origin of the overpotential for oxygen reduction[143]. They recognized that the high-spin ground state of O_2 is poorly described with DFT, leading to large errors of $\sim$70 kJ mol^{-1} in the water splitting reaction. To avoid this issue, they used the experimental reaction energy for water splitting $\frac{1}{2}O_2 + H_2 \longrightarrow H_2O$ (equal to $\Delta_f H_{H_2O}$) to avoid calculating O_2 with DFT, leading to the widely used "O_2 correction" that ensures the correct enthalpy of the water splitting reaction[143,168–170].

$$E_{O_2}^{\text{corr}} = 2\left(E_{H_2O} - E_{H_2} + \Delta_f H_{H_2O}\right) \tag{50}$$

where $\Delta_f H_{H_2O} = -238.9\,\text{kJ mol}^{-1}$ (ATcT[75]). Another widely used correction to DFT energies of gas-phase species for catalytic systems was proposed by Peterson et al.[144] for the electrochemical reduction of CO_2 on Cu(111). The authors applied a statistical sensitivity analysis to identify correction factors that align the reaction enthalpies from DFT with the experimental values. To determine these correction factors, they compared the DFT-derived reaction enthalpies of 21 hand-picked gas-phase reactions involving CO_2, CO, H_2, and H_2O with experimental values from the NIST database. Perturbing the energy of CO_2 and all species with an OCO backbone by 43.4 kJ mol^{-1} for the RPBE functional led to a minimum deviation between the experimental and DFT values. However, the agreement with all test reactions is not exact, and a discrepancy of $\approx$6 kJ mol^{-1} remained. In the same study, very different correction factors were necessary when using the PBE functional, highlighting the functional dependence of these corrections. In a follow-up study, Studt et al.[171] repeated the same analysis using the BEEF-vdW functional in GPAW, and they obtained the closest agreement when perturbing the CO_2 as well as the H_2 energy by 31.8 kJ mol^{-1} and 8.7 kJ mol^{-1}. In a different study by Studt et al.[172], a slightly different correction value for CO_2 of 39.6 kJ mol^{-1} was obtained using QuantumEspresso while the H_2 correction remained 8.7 kJ mol^{-1}. This discrepancy in correction factors highlights the slight differences between different electronic structure codes and numerical settings, making it challenging to generalize correction factors. Following these pioneering studies, a range of corrections for the gas-phase energies were proposed in the literature for nitrogen-containing species[60,168], O_2 [173], CO, CO_2 and other species depending on the xc-functional[77,174–176]. Urrego-Ortiz et al.[177] reviewed the gas-phase errors and corrections of DFT calculations for computational catalysis.

Fig. 8 shows the enthalpy diagram with the applied corrections to CO_2 and H_2 of 31.8 kJ mol^{-1} and 8.7 kJ mol^{-1} for the BEEF-vdW functional[171], respectively. The derivation of the $\Delta_A H$ follows the equations described in Section 4.2.2, with corrections added to the DFT gas-phase energies of CO_2 and H_2O. Notably, the reaction enthalpies will be the same regardless of whether corrections are added before or after anchoring the thermochemical network; however, if corrected species are used as anchors, then the correction will be distributed throughout the network, and the individual $\Delta_A H$ will depend on whether the correction was applied before or after anchoring. Thus, it is recommended that corrected species should not be used as anchors to avoid ambiguity about which species have been "corrected". Using these standard corrections, the overall reaction enthalpy with the gas-phase corrections is 159.4 kJ mol^{-1}, which still deviates from the experimental value by 10 kJ mol^{-1}. This error leads to an equilibrium constant that deviates by a factor of $\sim$4 from the exact value (see Fig. 7b). Compared to the "non-corrected" free energy diagram constructed directly from DFT energies, the only change is for the adsorption enthalpy of CO_2, which is increased by 31.8 kJ mol^{-1}, making CO_2* more stable. Since H_2 desorption does not appear in this reaction mechanism, the H_2 correction cancels out, and all other DFT reaction enthalpies are conserved. In general, the approach changes only the reaction energies for the adsorption/desorption steps where corrected molecules are involved.

Instead of applying corrections to specific molecules, it is possible to determine correction factors for bond types. The OCO correction of Peterson *et al.*[144] is a simple example, but more sophisticated approaches are possible. The usage of bond-additivity corrections to improve the accuracy of low level of theory methods is a well-established procedure in the gas-phase commu-

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

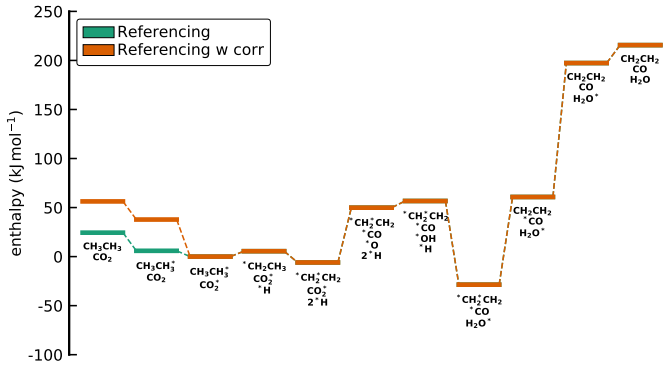

Fig. 8 Enthalpy diagram with the relative enthalpies derived from the corrected DFT energies of CO_2 ($+31.8\,kJ\,mol^{-1}$) and H_2 ($+8.7\,kJ\,mol^{-1}$) using referencing approach compared with the $\Delta_A H$ from the uncorrected DFT energies. The enthalpy diagrams are aligned to C_2H_6* + CO_2* to demonstrate that the adsorption enthalpy of CO_2 was corrected.

nity[167,178]. Bond-additivity corrections are increasingly used in catalysis. Christensen et al.[179] et al. identified that the error is not associated with the OCO backbone but rather with the C=O bond. They obtained a slightly better agreement with the experimental reaction enthalpies when applying corrections of $9.7\,kJ\,mol^{-1}$ to H_2 and $14.5\,kJ\,mol^{-1}$ for the C=O bond. Additional bond-additivity correction factors were obtained for a range of functionals by Granda-Marulanda et al.[77] for CO_2, CO, $-CH_x$, $-OH$, C=O, and (C=O)O. These fragments were hand-picked by experts, which requires chemical intuition and experience. Alternatively, it is possible to use machine-learning techniques like LASSO regression to determine the subgraphs that lead to the best results[165,180]. Overall, these correction schemes provide chemically intuitive ways to understand the corrections, and in the simplest cases are very straightforward to apply. However, the corrections do not guarantee any specific accuracy of gas-phase reactions and can become rather complex when generalized between different xc functionals.

4.3.2 Deriving enthalpies of formation of adsorbates using gas-phase reference species

It is possible to combine the $\Delta_f H$ of the adsorbates with the known global thermochemical network of gas-phase reactions to achieve reaction enthalpies of gas-phase reactions that are in exact agreement with the known $\Delta_f H$ from the ATcT. The approach to reference the DFT energies to the ATcT is similar to the approach described in Section 4.2.2. Assuming an atomic reference basis set with CH_4, H_2O, H_2, and Pt(111), we can again create a reaction to form the target from the references.

$$\nu_{CH_4} CH_4 + \nu_{H_2O} H_2O + \nu_{H_2} H_2 + \nu_* * \longrightarrow P$$

In Section 4.2.2, we assumed that these species are our anchors from which all other species are formed. Using the DFT energies of these species as the anchor values, it followed that the $\Delta_A H$ of CH_4, H_2O, H_2, and Pt(111) in the anchor frame A are 0 because it is a null reaction. In the global thermochemical network, $\Delta_f H$ of these species are referenced to the IUPAC anchor species in their standard state. Thus, $\Delta_f H$ of the reference species have known values as listed in Table 2. To utilize the tabulated $\Delta_f H$ of the gas-phase species from the global network, we have to reference the DFT data to the IUPAC anchors, which we do through our chosen reference species in the atomic basis set, shown in the thermochemical cycle in Fig. 9.

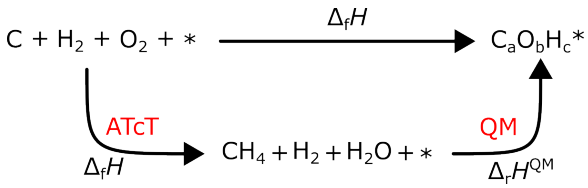

Fig. 9 Thermodynamic cycle to determine the $\Delta_f H$ of the species connected to the global thermochemical network.

From the thermochemical cycle, it can be easily seen that $\Delta_f H_P$ in the IUPAC anchor frame is the sum of the $\Delta_f H$ of the reference species and the reaction enthalpy to form the target from the references.

$$\Delta_f H_{P^*} = \Delta_r H^{QM} - \sum_{i \neq P}^{N} \nu_i \Delta_f H_i \tag{51}$$

$$\underbrace{\Delta_f H_{P^*}}_{\text{target}} = \underbrace{E_{P^*}}_{\text{target}} + \underbrace{\sum_{i \neq P}^{N} \nu_i E_i}_{\text{references}} - \underbrace{\sum_{i \neq P}^{N} \nu_i \Delta_f H_i}_{\text{references}} \tag{52}$$

We can also formulate this in a linear algebra notation leading to the following equation:

$$\underline{\mathbf{H}}_f = \underline{\mathbf{E}} + \underline{\underline{\mathbf{M}}}\,\underline{\mathbf{E}}^R - \underline{\underline{\mathbf{M}}}\,\underline{\mathbf{H}}_f^R \tag{53}$$

$$\underline{\mathbf{H}}_f = \underline{\mathbf{E}} + \underline{\underline{\mathbf{M}}} \left(\underline{\mathbf{E}}^R - \underline{\mathbf{H}}_f^R \right) \tag{54}$$

Including the slab into the matrix of stoichiometric coefficients leads to an equation that is similar to Eq. (42), with an additional term $-\underline{\underline{\mathbf{M}}}\,\underline{\mathbf{H}}_f^R$, which references enthalpies of formation to the ATcT thermochemical network. As shown previously in Section 4.2.2, $\underline{\underline{\mathbf{M}}}$ can conveniently be determined from $\underline{\mathbf{N}}$ and $\underline{\mathbf{N}}^R$ via Eq. (44). An alternative derivation of this approach was developed by Blöndal et al.[181], which relies on an adsorption reaction of the target species.

$$P + * \rightleftharpoons P^*$$

where the adsorption enthalpy is derived from the DFT energies and $\Delta_f H_* = 0\,kJ\,mol^{-1}$ by assertion as discussed in Section 4.1. Thus, the only unknown to solve is $\Delta_f H$ of the gas-phase precursor of the adsorbate. Many gas-phase precursors are unstable and, therefore, not tabulated in the ATcT. An *isogyric* reaction[90,182] to form the species from CH_4, H_2O, and H_2 is used to estimate $\Delta_f H$ of the precursor, which is similar to the ANL-0 approach[90]. $\Delta_f H$ of the gas-phase precursor also provides the reference to an existing thermochemical network as illustrated in Fig. 10. We derive the approach in detail in the SI and show how it is mathematically equivalent to Eq. (54).

The method depends on the DFT energies of the reference species, in this case CH_4, H_2O, and H_2. The approach is typi-

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

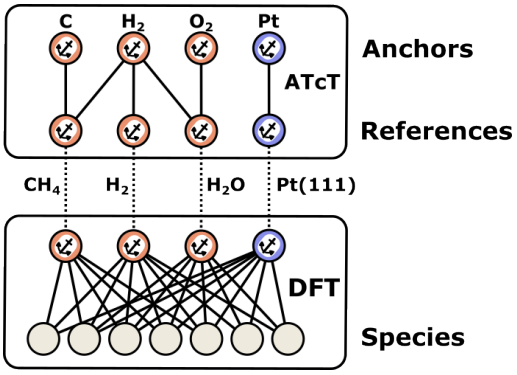

Fig. 10 Schematic illustration of referencing the DFT data to the global thermochemical network (here ATcT) using gas-phase reference species.

cally used with this reference species set since it is equivalent to an *isogyric* reaction[90,182]. However, other reference sets could be selected, in which case the final $\Delta_f H$ of the resulting reaction network and the resulting adsorption enthalpies will vary slightly depending on the selected reference species. For example, it is possible to use the least squares anchor set introduced in Section 4.2.3 to combine data from two thermochemical networks without explicitly selecting molecular references. In this case, all adsorption/desorption energies would be corrected by some amount, but on average, the squared difference between the DFT-derived and ATcT enthalpies of formation would be minimized.

Referencing the DFT energies to the existing ATcT database through the reference species allows us to integrate the available accurate $\Delta_f H$ of gas-phase species, which always ensures thermodynamically consistent mechanism. By using the highly accurate $\Delta_f H$ of the gas-phase species from the ATcT instead of the DFT values, we implicitly correct all DFT-derived adsorption/desorption reaction energies except for those of the reference species (see Fig. 11). The adsorption enthalpies of the references match the DFT energies (see SI). All reaction enthalpies of the surface reactions from DFT are still conserved and identical to those of the referencing approach. This method also reduces the computational workload, as it is no longer necessary to compute the energies of gas-phase species other than the references.

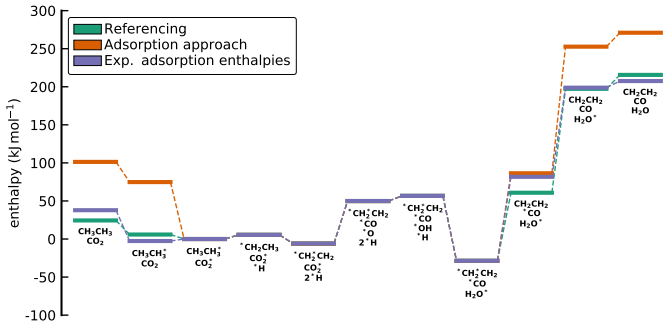

Fig. 11 Enthalpy diagram with the enthalpies of formation derived with the standard referencing approach, using the gas-phase $\Delta_f H$ as references, and using experimental enthalpies of formation of adsorbates as reference. The enthalpy diagrams are aligned to $C_2H_6^* + CO_2^*$ to demonstrate the changes in the adsorption enthalpies, while all surface reaction enthalpies are conserved.

This approach has been widely employed by Vlachos and coworkers[183–186] to convert DFT energies to $\Delta_f H$ referenced to the global thermochemical network. Vorotnikov et al.[183] used the method to construct a database of adsorbates for furan chemistry on Pd(111). Further, they used Eq. (52) to determine the $\Delta_f H$ of all gas-phase species (that are not the references) with the G4 level of theory rather than using ATcT values. This application illustrates an alternative use of the approach where gas-phase species and adsorbates are computed with different levels of theory but can be combined in a single thermochemical network since both are referenced to the same network. Integrating the accurate $\Delta_f H$ of gas-phase species and the $\Delta_f H$ of adsorbates derived from DFT in a global thermochemical network is crucial when building coupled homogeneous/heterogeneous reaction mechanisms[181] or for open-ended mechanism exploration with automated mechanism generation software[68,70,187–189]. Some of the authors have used the method for the microkinetic modeling of the transient CO_2 methanation on Ni(111)[71] or the temperature-programmed desorption of CO_2[190].

It is also possible to derive elemental chemical potentials from this approach that are conveniently referenced to the global thermochemical network. Using the actual $\Delta_f H$ of the reference species basically results in the correction of the DFT-derived elemental chemical potentials to match the experimental $\Delta_f H$ of the reference species via

$$\mu_C = \left(E_{CH_4} - 2E_{H_2}\right) + \left(\Delta_f H_{CH_4} - 2\Delta_f H_{H_2}\right) \quad (55)$$

$$\mu_O = \left(E_{H_2O} + E_{H_2}\right) + \left(\Delta_f H_{H_2O} - \Delta_f H_{H_2}\right) \quad (56)$$

$$\mu_H = 0.5E_{H_2} + 0.5\Delta_f H_{H_2} \quad (57)$$

$$\mu_* = E_{Pt(111)} + \Delta_f H_{Pt(111)} \quad (58)$$

With these corrected elemental chemical potentials, the method described in Section 4.2.1 can be used to derive $\Delta_f H$ of the adsorbates, which can the be combined with accurate $\Delta_f H$ of gas-phase species from the ATcT to achieve thermodynamic consistency.

4.3.3 Deriving enthalpies of formation of adsorbate connected to the global thermochemical network through adsorbates

Blaylock et al.[191,192] developed a similar approach, where they also used an expression similar to Eq. (54). Instead of gas-phase reference species, they used adsorbates as reference species for the atomic reference basis set. They determined $\Delta_f H$ of *H, *O, and *CO on Ni(111) from experimentally measured adsorption enthalpies[193–195]. In combination with the known $\Delta_f H$ of the gas-phase precursors (H_2, O_2, CO) from a global thermochemical network like ATcT, we can derive $\Delta_f H$ of the reference adsorbates using the thermochemical cycle in Fig. 2. $\Delta_f H$ of the adsorbed reference species are referenced to the global thermochemical network, and all other adsorbates are then referenced to these reference species according to

$$\nu_{*H}{}^*H + \nu_{*O}{}^*O + \nu_{*CO}{}^*CO + \nu_{**} \longrightarrow P^* \quad (59)$$

which is illustrated in the thermochemical cycle in Fig. 12

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

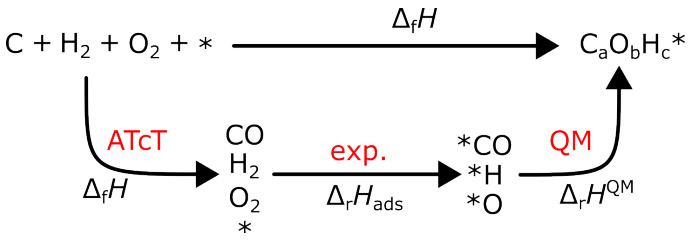

Fig. 12 Thermodynamic cycle for the connection of the DFT energies to the global thermochemical network via experimentally determined enthalpies of adsorption for a reference basis set.

The method to determine $\Delta_f H_{P^*}$ is straightforward and similar to the method in Section 4.3.2. $\Delta_f H_{P^*}$ is calculated from the reaction enthalpy to form the target from the references from DFT and the sum of the $\Delta_f H$ of the references.

$$\Delta_f H_{P^*} = \Delta_r H^{QM} - \sum_{i \neq P}^{N} \nu_i \Delta_f H_i \tag{60}$$

$$\underbrace{\Delta_f H_{P^*}}_{\text{target}} = \underbrace{E_{P^*}}_{\text{target}} + \underbrace{\sum_{i \neq P}^{N} \nu_i E_i}_{\text{references}} - \underbrace{\sum_{i \neq P}^{N} \nu_i \Delta_f H_i}_{\text{references}} \tag{61}$$

In linear algebra form this reads as

$$\underline{\mathbf{H}}_f = \underline{\mathbf{E}} + \underline{\underline{\mathbf{M}}} \left(\underline{\mathbf{E}}^R - \underline{\mathbf{H}}_f^R \right) \tag{62}$$

This is mathematically identical to the prior approach (Eq. (54)), with the difference being that the enthalpies of adsorbed species, rather than gas-phase species, are used as references. For the construction of the stoichiometry matrix $\underline{\underline{\mathbf{M}}}$, we refer the reader to Section 4.2.2. Since the approach connects the DFT data with the ATcT, it is possible to replace all DFT energies of gas-phase species with accurate $\Delta_f H$ form the ATcT to achieve thermodynamic consistency. The resulting enthalpy diagram using this approach is shown in Fig. 12. Similar to the previous method, this approach does not affect the reaction enthalpies of surface steps since it uses an atomic reference basis set. However, all adsorption/desorption reaction enthalpies are changed except for the reference species. In the previous approach, the adsorption enthalpies of CH_4, H_2, and H_2O were fixed at the DFT values. The difference in this approach is that the adsorption enthalpies of CO, H_2, and O_2 correspond now to experimental values (see Fig. 13). This adjustment leads to noticeable changes in the adsorption enthalpies of some species. The biggest change occurs for the adsorption of CO_2, which changes from strongly exothermic to mildly endothermic. Typically, CO_2 binds only weakly through physisorption on the Pt(111) surface[196]. Using gas-phase reference $\Delta_f H$ indicates that the CO_2 binds strongly to the surface in disagreement with experiments. Mhadeshwar et al.[123] also describe the use of the experimental $\Delta_f H$ of adsorbates as a basis set to reference the DFT energies to the global thermochemical network, thereby ensuring thermodynamic consistency of the reaction mechanism with known adsorption enthalpies.

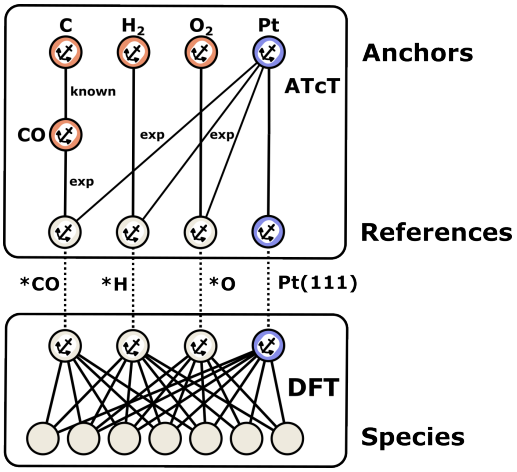

Fig. 13 Connecting the DFT data with the global thermochemical network (here ATcT) by referencing DFT data to experimental $\Delta_f H$ of adsorbates.

4.4 Indirect correction schemes with molecular fragment reference basis sets

The methods that have been introduced in prior sections rely on an atomic basis set, with one reference species defined per element. Depending on the scheme, these references may be gas-phase or adsorbed species, but the approaches presented thus far effectively act as adsorption energy corrections. They ensure that gas-phase reaction energies are consistent with ATcT, and that surface reactions are consistent with DFT, while adsorption energies are adjusted in various ways to ensure this is the case. Each method makes different assumptions about how to adjust the adsorption energies. However, since surface reaction enthalpies are rarely known, the most common approach is to assume that DFT surface reaction enthalpies are accurate. The logic for this is that error cancellation in DFT is likely the largest for surface reactions since both products and reactants have similar electronic environments (adsorbates on surfaces).

While there is certainly a higher degree of error cancellation for pure surface elementary steps than adsorption reactions, error cancellation is not automatically guaranteed, and the degree of cancellation can vary widely depending on the nature of the adsorbed species. To demonstrate, we show the computed reaction enthalpy for a set of surface reactions using the BEEF-vdW and RPBE functional compared to the experimental values in Fig. 14. The experimental $\Delta_f H$ of the adsorbates were derived from the adsorption enthalpies reported by Silbaugh and Campbell[95]. While most of the predicted reaction enthalpies are within $\pm 25\,\mathrm{kJ\,mol^{-1}}$, there are deviations of up to $90\,\mathrm{kJ\,mol^{-1}}$ from the experiment. Additionally, results from the different functionals can vary substantially. For example, the reaction enthalpy for the dissociation of *OH determined from the RPBE xc-functional agrees within chemical accuracy with the experiments, while the BEEF-vdW value deviates by $35\,\mathrm{kJ\,mol^{-1}}$. This large deviation and sensitivity to functional choice indicates a low degree of error cancellation. There are also reactions for which both functionals give similar results, indicating a higher degree of error cancellation since the specific errors of each xc functional are canceled out.

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Consequently, the generalization of accurate reaction enthalpies from DFT for all surface elementary reactions due to error cancellation is not possible.

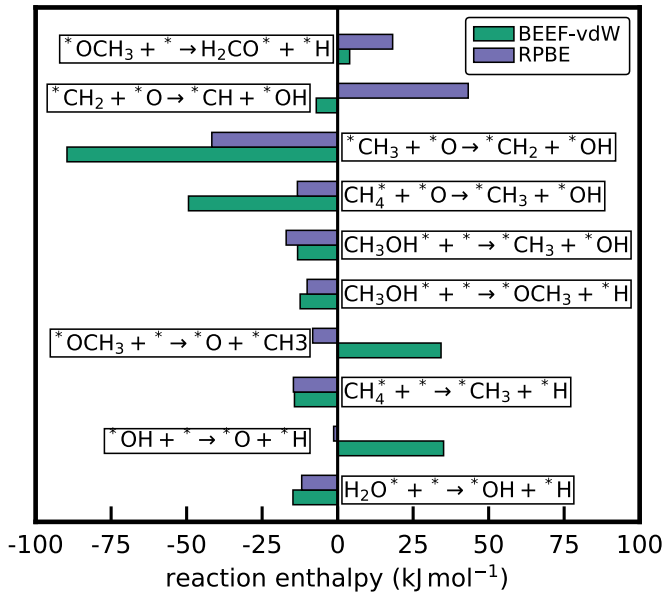

Fig. 14 Comparison of predicted reaction enthalpies from the BEEF-vdW and RPBE functional compared to the experimental value for a set of surface reactions.

It is possible to maximize the degree of error cancellation by constructing reactions that conserve the bonding environment and hybridization on both sides of the reaction. This approach moves away from atomic reference sets, instead selecting reference reactions by maximizing the similarity of molecular fragments. The gas-phase community has used these error-cancellation reactions to derive $\Delta_f H$ of species with chemical accuracy at the DFT level of theory[197,198]. Instead of conserving the reaction enthalpies of surface reactions that are relevant to a given chemical reaction network, the goal is to conserve the reaction enthalpies for reactions with the highest degree of error cancellation. The key is to construct reactions that conserve the molecular structure and bonding environment of the target molecule to maximize error cancellation[182,199–201]. Raghavachari and co-workers[199,200] developed the connectivity-based hierarchy (CBH) that automatically constructs error cancellation reactions for a gas-phase target. The order of the CBH approach is conservation of spin pairs (isogyric, CBH-0), bond types (isodesmic reaction, CBH-1), immediate connectivity of the heavy atoms (hypohomodesmotic, CBH-2), or immediate connectivity of the bonding environment (hyperhomodesmotic, CBH-3). Climbing this ladder leads to increasing error cancellation and more accurate $\Delta_f H$. Fig. 15 shows the order of the error cancellation reactions with a gas-phase example for 1-propyl.

To illustrate this concept, we calculate gas-phase reaction enthalpies at various rungs of the CBH ladder using BEEF-vdW, illustrated in Fig. 15. The simplest approach to calculate the $\Delta_f H$ is via the total atomization energy from the BEEF-vdW energies. This approach results in a $\Delta_f H$ that deviates wildly from

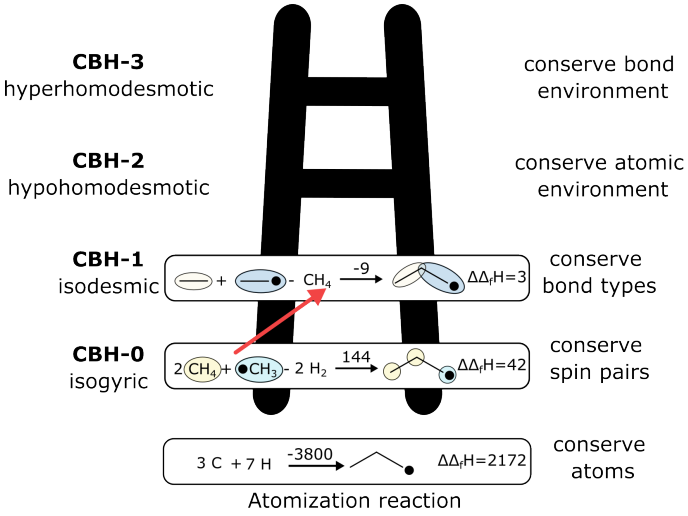

Fig. 15 Types of error cancellation reactions in the CBH[199,200] scheme to determine the enthalpy of formation of the 1-propyl radical using BEEF-vdW DFT energies. The reported differences in $\Delta_f H$ are with respect to the ATcT value of 118 kJ mol^{-1}. The value above the arrow is the reaction enthalpy and all values are reported in kJ mol^{-1}. The atomization reaction is not part of the CBH method.

the experimental value, with an error of 2,172 kJ mol^{-1}. This error arises due to the implicit assumption that the atomization energies are correctly computed with BEEF-vdW, but it is well known that much higher levels of theory are needed to accurately compute total atomization energies[86,199–202]. The isogyric reaction gives a much better agreement with the experiment, with a deviation of only 42 kJ mol^{-1}, but the deviation is still large enough to cause equilibrium constants to be off by orders of magnitude. The highest possible rung in the CBH scheme for this reaction is an isodesmic reaction for 1-propyl. With this reaction, $\Delta_f H$ is determined *to within chemical accuracy (3 kJ mol^{-1}) of the accurate ATcT value from BEEF-vdW energies*. Although the gas-phase community has highly accurate electronic structure methods, error cancellation is still a very active field of research[202,203]. This example illustrates the power of error cancellation in a gas-phase context, and the same concepts hold for surface reactions, although some modifications are required.

4.4.1 Deriving enthalpies of formation of adsorbates using adsorbate reference species in isodesmic reactions

Recently, Kreitz et al.[128] introduced the concept of error cancellation reactions for adsorbates by extending the CBH approach. The CBH approach provides a convenient way to connect the DFT data with the global thermochemical network by using accurate and independent reference $\Delta_f H$ for the adsorbates that serve as the bond-type fragments used in the reference reactions. This necessity is a limitation for many catalytic surfaces where experimental or high-level theory data is not available. However, for Pt(111) (and some other close-packed facets), there is enough experimental data available in the literature, which is summarized by Silbaugh and Campbell[95], to decompose all $C_xH_yO_z$ adsorbates into bond types using isodesmic reactions (CBH-1). It is not possible to use the lower CBH-0 rung due to inconsistencies with

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

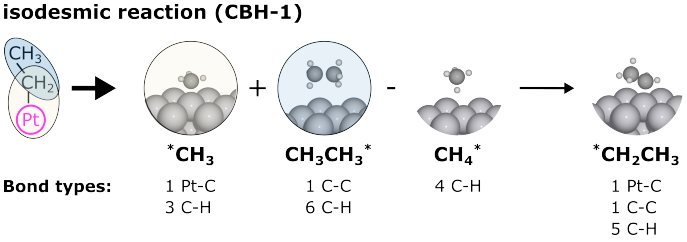

Fig. 16 Isodesmic reaction to determine the $\Delta_f H$ of adsorbed ethyl ($^*CH_2CH_3$) using the CBH approach for adsorbates developed by Kreitz et al.[128]. This illustration shows that the C to which ethyl binds to the surface is counted twice, which is why a CH_4 needs to be subtracted.

the conservation of surface sites, which conveniently cancel out at the CBH-1 rung, as explained in detail in ref. 128. The $\Delta_f H$ of the adsorbates used as the reference fragments are referenced to the ATcT and all other adsorbates are referenced to the set of reference adsorbates. Similarly to Section 4.3.3, we avoid the use of a gas-phase precursors and combine DFT with experimental adsorbate $\Delta_f H$ directly. Table 3 contains a list of the $\Delta_f H$ of all the species that were used for the bond types and additional experimental values for the case study. In the literature, these experimental values are only used to benchmark the DFT calculations; here, we show how to combine them with DFT values so that they can all be included in a consistent thermochemical network. $\Delta_f H$ of the target is determined from an isodesmic reaction using the fragments F, where F are adsorbates. An example of the isodesmic reaction for adsorbed ethyl ($^*CH_2CH_3$) is given in Fig. 16.

bond type	species	$\Delta_f H(0K)$	ref.
C–O	CH_3OH^*	-245.0	122
C=C	$CH_2CH_2^*$	22.1	204
C=O	H_2CO^*	-159.3	205
C–C	$CH_3CH_3^*$	-96.0	67
C–H	CH_4^*	-81.3	67
O–H	H_2O^*	-267.9	67
Pt–C	$^*CH_3$	-47.2	121
Pt≡C	$^*CH$	-35.8	206
Pt=C	$^*CH_2$[a]	46.5	206
Pt–O	$^*OH$[b]	-164.7	207
Pt–H	$^*H$	-32.7	67
Pt=O	$^*O$	-103.7	67
Pt=C=O	$^*CO$	-230.9	67

Table 3 $\Delta_f H$ at 0 K of the species that were used as fragments for the bond types to construct the isodesmic reactions. All enthalpies are in $kJ\,mol^{-1}$. The table is adapted from Kreitz et al.[128] and details for the derivation of the $\Delta_f H$ can be found there.

The various bond types in this reaction are summarized in the table for demonstration purposes and show that all bond types are perfectly balanced on both sides of the reaction. An open-source software tool is available to construct these structure-preserving reactions automatically for adsorbates[128,208]. The method for calculating the $\Delta_f H^*_{CH_2CH_3}$ from this reference reaction is similar to the approaches described before. First, the zero-point corrected DFT energies are used to compute the reaction enthalpy.

$$\Delta_r H^{QM} = \underbrace{E_P}_{\text{target}} + \underbrace{\sum_{i\neq P}^{N} \nu_i E_i}_{\text{references}} \tag{63}$$

$$\Delta_r H^{QM} = E^*_{CH_2CH_3} - E^*_{CH_3} - E_{C_2H_6^*} + E_{CH_4^*} \tag{64}$$

Assuming that $\Delta_r H^{QM} = \Delta_r H$, we can calculate $\Delta_f H$ of the target with the known $\Delta_f H$ of the reference adsorbates.

$$\underbrace{\Delta_f H_P}_{\text{target}} = \Delta_r H - \underbrace{\sum_{i\neq P}^{N} \nu_i \Delta_f H_i}_{\text{references}} \tag{65}$$

$$\Delta_f H^*_{CH_2CH_3} = \Delta_r H^{QM} + \Delta_f H_{CH_3CH_3^*} + \Delta_f H^*_{CH_3} - \Delta_f H_{CH_4^*} \tag{66}$$

The missing additional species for which isodesmic reactions can be used from our case study are $^*CH_2{}^*CH_2$ and $^*CO_2$.

$$CH_3CH_3^* + 2{}^*CH_3 - 2CH_4^* \longrightarrow {}^*CH_2{}^*CH_2$$

$$2H_2CO^* - CH_4^* \longrightarrow CO_2^*$$

The linear algebra notation for this approach is straightforward and identical to Eq. (54),

$$\underline{\mathbf{H}}_f = \underline{\mathbf{E}} + \underline{\underline{\mathbf{M}}}\,\underline{\mathbf{E}}^R - \underline{\underline{\mathbf{M}}}\,\underline{\mathbf{H}}_f^R \tag{67}$$

$$\underline{\mathbf{H}}_f = \underline{\mathbf{E}} + \underline{\underline{\mathbf{M}}}\left(\underline{\mathbf{E}}^R - \underline{\mathbf{H}}_f^R\right) \tag{68}$$

where $\underline{\underline{\mathbf{M}}}$ is the matrix of stoichiometric coefficients for the isodesmic reactions, which can be enumerated using the algorithm or software described in ref. 128. The stoichiometric matrix can again be derived using linear algebra, but in this case, from a matrix $\underline{\mathbf{F}}$ (number of molecules × number of fragments) containing the composition of fragments of the target.

$$\underline{\underline{\mathbf{F}}} = \begin{array}{c} \text{Fragments} \\ \begin{array}{cccc} \text{C–O} & \text{C=O} & \text{C–C} & \dots \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & \dots \\ 0 & 0 & 1 & \dots \\ 0 & 1 & 0 & \dots \end{bmatrix} \end{array} \begin{array}{c} \\ \begin{array}{c} {}^*CH_2H_3 \\ {}^*CH_2{}^*CH_2 \\ CO_2^* \end{array} \end{array} \quad \text{Species} \tag{69}$$

Analogously, we can define a matrix with the fragments of the reference species $\underline{\mathbf{F}}^R$.

$$\underline{\underline{\mathbf{F}}}^R = \begin{array}{c} \text{Fragments} \\ \begin{array}{cccc} \text{C–O} & \text{C=O} & \text{C–C} & \dots \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \end{array} \begin{array}{c} \\ \begin{array}{c} CH_3OH^* \\ H_2CO^* \\ CH_3CH_3^* \\ \vdots \end{array} \end{array} \quad \text{Reference} \tag{70}$$

The stoichiometry matrix is then calculated via

$$\underline{\underline{\mathbf{M}}} = -\underline{\underline{\mathbf{F}}}\,\underline{\underline{\mathbf{F}}}^{R^{-1}} \tag{71}$$

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

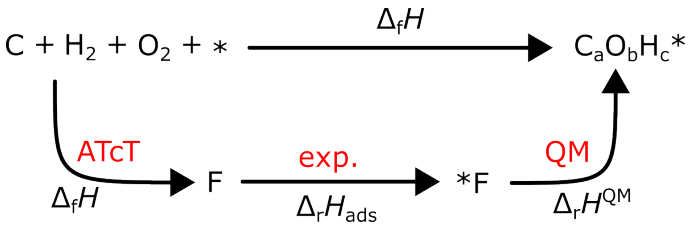

Fig. 17 Thermochemical cycle for the isodesmic reactions.

which can be compared to Eq. (43) to highlight the use of a "fragment" basis rather than an "atomic" basis in this approach.

The CBH methodology for adsorbates can only be applied for molecules that can be broken down into isodesmic reactions. Thus, the method cannot be used for small adsorbates like *O, *H, or *CO. In this work, we can use experimental $\Delta_f H$ for the parameterization of the microkinetic model (see Table 3). Alternatively, it is possible to perform higher-level of theory calculations for these adsorbates to determine accurate $\Delta_f H$. The CBH approach integrates available experimental data for the reference adsorbates, accurate gas-phase species from the ATcT, and DFT energies into a global thermochemical network as shown in Fig. 17. Additionally, systematic error cancellation provides a more accurate $\Delta_f H$ of adsorbates from DFT. Kreitz et al.[128] obtained a $\Delta_f H$ for $^*OCH_3$ and $C_3H_8^*$ that is within chemical accuracy of the experimental value using BEEF-vdW DFT energies. However, isodesmic reactions do not necessarily have the maximum degree of error cancellation[199,201], *e.g.*, the degree of error cancellation for CO_2^* is rather low. Conserving only the bond types does not always conserve the correct hybridization and the molecular orbitals. A good indicator for the effectiveness of the error cancellation is the reaction enthalpy of Eq. (63)[208]. High reaction enthalpies indicate a low degree of error cancellation, and climbing to the next highest rung would be preferable. Unfortunately, this depends on the size of the target adsorbate. Additionally, independent and accurate reference $\Delta_f H$ are needed to populate the fragments for the hypohomodesmotic level. Yet, the approach provides a way to leverage a relatively small number of strategically selected highly accurate reaction enthalpies from experiment or wavefunction theories and use these to elevate the accuracy of other reaction enthalpies by maximizing cancellation of error.

While the molecular fragment reference basis sets still conserve the stoichiometry, they also conserve the bond types in the molecule, which leads to differences between the reference fragments in the products and reactants. Consequently, these approaches do not necessarily conserve the surface reaction enthalpies of a mechanism at the DFT values. Instead, they conserve surface reaction enthalpies of alternative reactions that maximize error cancellation for a given electronic structure method[182,197–199].

Fig. 18 shows the enthalpy diagram with $\Delta_f H$ from experiments, ATcT, and the isodesmic reactions. Since the enthalpies of the gas-phase species are directly from the ATcT, we match the reaction enthalpy exactly, which is identical to the adsorption reaction approach in Section 4.3.2. The DFT energy of the empty slab is also no longer needed. The reaction enthalpies now vary considerably in comparison to the previous methods. The dissociation of ethane to ethyl changed from endothermic to exothermic and the subsequent dissociation step is even more exothermic. Furthermore, the formation of *OH is also exothermic, while the endothermicity for the dissociation of CO_2^* increased significantly. Assuming the experimentally measured adsorption enthalpies are more accurate than DFT, it follows that the free energy pathway constructed from the CBH expansion approach should also be more accurate. However, it is only possible to calculate $\Delta_f H$ using the CBH approach in cases where the experimental (or other highly accurate) reaction enthalpies of the requisite reference species are known.

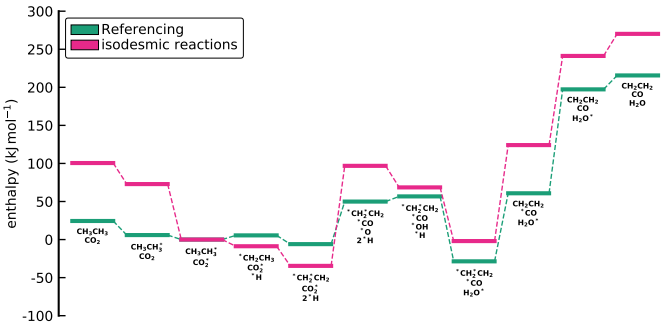

Fig. 18 Enthalpy diagram with enthalpies of formation derived with the standard referencing approach and using the molecular fragment basis set with isodesmic reactions. The enthalpy diagrams are aligned to $C_2H_6^* + CO_2^*$ to demonstrate that using a molecular fragment basis set changes the adsorption enthalpies and the surface reaction enthalpies.

5 Converting between reaction enthalpies and formation enthalpies

The approaches in Sections 4.2.1 to 4.2.3 have shown that the reaction enthalpies calculated from the total DFT energies are conserved, regardless of the choice of atomic reference species. We can use this fact and calculate $\Delta_A H$ if only reaction enthalpies are reported for a mechanism. This problem cannot be formulated for a single species, and it directly becomes a linear algebra problem. For a chemical kinetics network involving n chemical species, a total of $n_a + n_s$ reference species must be defined, where n_a is the number of unique atomic elements of which all species consist and n_s is the number of surface sites (typically one). Let $\underline{\mathbf{E}}$ be the zero-point corrected energies of chemical species, $\underline{\mathbf{H}}_A$ be the enthalpies of formation of the same species given defined references A and $\underline{\mathbf{H}}_r$ be the reaction energies according to the stoichiometry matrix $\underline{\underline{\mathbf{M}}}$. The relationship between $\underline{\mathbf{H}}_r$ and $\underline{\mathbf{E}}$ is a simple linear combination, shown in Eq. (72).

$$\underline{\mathbf{H}}_r = \underline{\underline{\mathbf{M}}}^\intercal \underline{\mathbf{H}}_A = \underline{\underline{\mathbf{M}}}^\intercal \underline{\mathbf{E}} \tag{72}$$

Since the number of elementary reactions is typically greater than or equal to the number of chemical species, *i.e.*, $m \geq n$, there exists a direct mapping between $\underline{\mathbf{E}}$ and $\underline{\mathbf{H}}_r$. Yet, no obvious mapping exists in the opposite direction since this is an under-constrained linear system of equations. The Moore-

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Penrose[209–211] pseudo-inverse of the stoichiometry matrix $\underline{\underline{\mathbf{M}}}^{+}$ can be used to construct the inverse mapping. To solve this inverse mapping, it is first necessary to select anchor species for each element from the mechanism, *e.g.*, $A = [C_2H_6, CO, H_2O, Pt(111)]$. We can separate the stoichiometry matrix $\underline{\underline{\mathbf{M}}}$ into the anchor species $\underline{\underline{\mathbf{M}}}^{\mathrm{A}}$ and non-anchor species $\underline{\underline{\hat{\mathbf{M}}}}$, leading to Eq. (73).

$$\underline{\mathbf{H}}_{\mathrm{r}} = \underline{\underline{\mathbf{M}}}^{\intercal}\underline{\mathbf{H}}_{\mathrm{A}} = \begin{bmatrix} \underline{\underline{\hat{\mathbf{M}}}} \\ \\ \underline{\underline{\mathbf{M}}}_{\mathrm{A}} \end{bmatrix}^{\intercal} \begin{bmatrix} \underline{\hat{\mathbf{H}}}_{\mathrm{A}} \\ \\ \underline{\tilde{\mathbf{H}}}_{\mathrm{A}} \end{bmatrix} = \underline{\underline{\hat{\mathbf{M}}}}^{\intercal}\,\underline{\hat{\mathbf{H}}}_{\mathrm{A}} + \underline{\underline{\mathbf{M}}}_{\mathrm{A}}^{\intercal}\,\underline{\tilde{\mathbf{H}}}_{\mathrm{A}} \tag{73}$$

In Eq. (73), we separate $\underline{\mathbf{H}}_{\mathrm{A}}$ into the anchor species $\underline{\tilde{\mathbf{H}}}_{\mathrm{A}}$ and the unknown $\underline{\hat{\mathbf{H}}}_{\mathrm{A}}$. All enthalpies are referenced to the DFT energies of the anchor species. Thus, the relative enthalpies of the anchor species $\underline{\tilde{\mathbf{H}}}_{\mathrm{A}}$ are 0. We can now rearrange the equation to compute the unknown $\underline{\hat{\mathbf{H}}}_{\mathrm{A}}$ from the reaction enthalpies referenced to the set of anchor species.

$$\underline{\underline{\hat{\mathbf{M}}}}^{\intercal}\underline{\hat{\mathbf{H}}}_{\mathrm{A}} = \underline{\mathbf{H}}_{\mathrm{r}} - \underline{\underline{\mathbf{M}}}_{\mathrm{A}}^{\intercal}\underline{\tilde{\mathbf{H}}}_{\mathrm{A}} \tag{74}$$

$$\underline{\hat{\mathbf{H}}}_{\mathrm{A}} = \left(\underline{\underline{\hat{\mathbf{M}}}}^{\intercal}\right)^{+}\left(\underline{\mathbf{H}}_{\mathrm{r}} - \underline{\underline{\mathbf{M}}}_{\mathrm{A}}^{\intercal}\underline{\tilde{\mathbf{H}}}_{\mathrm{A}}\right) \tag{75}$$

$$\underline{\hat{\mathbf{H}}}_{\mathrm{A}} = \left(\underline{\underline{\hat{\mathbf{M}}}}^{\intercal}\right)^{+}\left(\underline{\underline{\mathbf{M}}}^{\intercal}\,\underline{\mathbf{E}} - \underline{\underline{\mathbf{M}}}_{\mathrm{A}}^{\intercal}\underline{\tilde{\mathbf{H}}}_{\mathrm{A}}\right) \tag{76}$$

This general form allows intuitive change of references. It is possible to use this method to create a thermochemical network that is referenced to the ATcT by defining the enthalpies of formation of the reference species to have the actual experimental values (instead of using the DFT values, which is equivalent to assuming $\underline{\tilde{\mathbf{H}}}_{\mathrm{A}} = 0$). When using gas-phase reference species, this results in the same enthalpies of formation as the method in Section 4.3.2, and when using adsorbates, this is equivalent to Section 4.3.3. Integrating the DFT data into the global thermochemical network allows to replace gas-phase DFT energies with accurate $\Delta_f H$ from the ATcT. The interested reader is referred to the SI for a more detailed discussion of the linear algebra of this approach. This approach of calculating enthalpies of formation directly from reaction enthalpies can serve as a convenient framework for new methods of combining data from various sources or identifying discrepancies in existing databases.

6 Summary & Conclusions

This review paper presents a cohesive collection of thermochemical concepts and linear algebra methods for calculating key thermochemical quantities in computational catalysis. These methods provide tools to combine data from various sources, to integrate experimental and *ab initio* data, and to convert between different formats. We have defined a notation for evaluating the thermochemistry of adsorbates from DFT data that is consistent with the existing gas-phase thermochemical conventions. All available methods in the literature for converting a set of DFT data into enthalpies of formation were described and explained following the consistent notation. The summary of the methods in Table 4 highlights that they can all be reduced to a simple linear algebra equation to anchor/reference the DFT data and to create a local or global thermochemical network. The difference in the methods is the choice of the anchor/reference species and the source of their enthalpies of formation. This review focused only on the enthalpy of formation at 0 K as it is the foundation of all thermochemistry and has the biggest contribution to the Gibbs free energy. $\Delta_f H$ at 0 K can be converted to $\Delta_f H$ at other temperatures or to Gibbs free energies of formation by applying finite-temperature corrections through *e.g.*, entropic contributions.

The approaches can be classified according to their use of an atomic reference basis sets or molecular fragment reference basis sets. In an atomic reference basis set method, every element must be assigned a single reference species (typically a closed-shell gas-phase molecule), which can be abstracted into elemental chemical potentials. The atomic reference basis set methods always conserve the reaction enthalpies as if they were calculated from the plain DFT energies, so long as correction factors are not applied and data from different sources are not mixed together. Conserving all DFT reaction enthalpies does not reproduce the thermodynamics of the overall gas-phase reaction from the accurate gas-phase $\Delta_f H$ of global thermochemical networks like the ATcT. Therefore, correction factors are applied to DFT energies to reduce discrepancies by adjusting adsorption enthalpies. However, the only method to ensure an exact agreement is to integrate the DFT data into the global thermochemical networks by anchoring it to the gas-phase or adsorbates $\Delta_f H$ of the references.

Instead of atomic references, a fragment reference basis set using *e.g.* isodesmic reactions can be used to integrate data from various sources (DFT, accurate gas-phase $\Delta_f H$ from the ATcT, and experimental adsorption enthalpies) into a global thermochemical network. The isodesmic reference reactions of this basis set lead to an increased accuracy of $\Delta_f H$ of adsorbates by maximizing error cancellation. Finally, we show how the linear algebra framework can also be used to "invert" the problem of constructing reaction enthalpies from $\Delta_f H$. This approach allows direct calculation of $\Delta_f H$ from tabulated reaction enthalpies, providing an alternative perspective on how thermochemical data can be converted and combined.

In conclusion, these tools can help researchers make better use of existing DFT datasets and facilitate the storage of new DFT data, preventing unnecessary waste of computational resources. The usage of a standardized reference frame has led to the proliferation of accurate thermochemical data in the gas-phase community, with wide-ranging implications on better reaction mechanisms and accurate models. The goal for the catalysis community should be to establish similar standards and methods. It is necessary to align the adsorbate thermochemistry concepts with the gas-phase community to integrate accurate thermophysical data of gas-phase species and adsorbates into a single global thermochemical network like the ATcT. Through the integration in this global network, it will be possible to simply search for the enthalpy of formation of an adsorbate in a database that contains the most accurate thermochemical information.

Author Contributions

B.K. - conceptualization, funding acquisition, data curation, formal analysis, investigation, software, writing - original draft, writing - review & editing; G.S.G - conceptualization, data curation,

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Method	Equation	Reference set	References	Data	Network type	Comments
Chemical elemental potentials	$\underline{\mathbf{H}}_{\mathrm{A}} = \underline{\mathbf{E}} - \underline{\underline{\mathbf{N}}}\,\underline{\mu}$	atomic	Gas-phase species	DFT	local	Convenient framework for dealing with non-standard reaction conditions (T, p, environment, applied potential)
Regression of elemental potentials	$\underline{\mathbf{H}}_{\mathrm{LS}} = \underline{\mathbf{E}} - \underline{\underline{\mathbf{N}}}\,\underline{\mu}^{\mathrm{LS}}$	atomic	Entire mechanism	DFT	local	Aligning DFT data from different sources
DFT anchors	$\underline{\mathbf{H}}_{\mathrm{A}} = \underline{\mathbf{E}} + \underline{\underline{\mathbf{M}}}\,\underline{\mathbf{E}}^{\mathrm{A}}$	atomic	Gas-phase species	DFT	local	Easy to use
DFT anchors with corrections	$\underline{\mathbf{H}}_{\mathrm{A}} = \underline{\mathbf{E}} + \underline{\underline{\mathbf{M}}}\,\underline{\mathbf{E}}^{\mathrm{A}}$	atomic	Gas-phase species	DFT	local	Increased accuracy of overall reaction enthalpy
Gas-phase references	$\underline{\mathbf{H}}_{\mathrm{f}} = \underline{\mathbf{E}} + \underline{\underline{\mathbf{M}}}\,\left(\underline{\mathbf{E}}^{\mathrm{R}} - \underline{\mathbf{H}}_{\mathrm{f}}^{\mathrm{R}}\right)$	atomic	Gas-phase species	DFT, ATcT	global	Accurate gas-phase $\Delta_{\mathrm{f}}H$
Adsorbate references	$\underline{\mathbf{H}}_{\mathrm{f}} = \underline{\mathbf{E}} + \underline{\underline{\mathbf{M}}}\,\left(\underline{\mathbf{E}}^{\mathrm{R}} - \underline{\mathbf{H}}_{\mathrm{f}}^{\mathrm{R}}\right)$	atomic	Adsorbates	DFT, ATcT, Exp.	global	Accurate gas-phase $\Delta_{\mathrm{f}}H$ and 3 experimental H_{ads}
Isodesmic reactions	$\underline{\mathbf{H}}_{\mathrm{f}} = \underline{\mathbf{E}} + \underline{\underline{\mathbf{M}}}\,\left(\underline{\mathbf{E}}^{\mathrm{R}} - \underline{\mathbf{H}}_{\mathrm{f}}^{\mathrm{R}}\right)$	bond type	Adsorbates	DFT, ATcT, Exp.	global	Accurate gas-phase $\Delta_{\mathrm{f}}H$, many experimental H_{ads}, and error-cancellation for surface reactions
Converting $\Delta_{\mathrm{r}}H$ to $\Delta_{\mathrm{A}}H$	$\underline{\mathbf{H}}_{\mathrm{A}} = \left(\underline{\underline{\hat{\mathbf{M}}}}^{\mathsf{T}}\right)^{+}\left(\underline{\underline{\mathbf{M}}}^{\mathsf{T}}\,\underline{\mathbf{E}} - \underline{\underline{\mathbf{M}}}_{\mathrm{A}}^{\mathsf{T}}\,\tilde{\underline{\mathbf{H}}}_{\mathrm{A}}\right)$	atomic	Gas-phase species or adsorbates	DFT, (ATcT, Exp.)	local (or global)	Useful if only reaction enthalpies are reported

Table 4 Summary of all discussed methods using the consistent notation with additional information on the specifics of each method.

formal analysis, software, writing - original draft; S.J.S. - formal analysis, investigation, software; A.A.P. - funding acquisition, writing - review & editing; D.H.B - writing - review & editing; D. N. - writing - original draft, investigation; C.F.G - conceptualization, funding acquisition, project administration, resources, supervision, writing - review & editing; A.J.M - conceptualization, funding acquisition, investigation, methodology, project administration, resources, supervision, writing - original draft, writing - review & editing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

BK, AAP, CFG, and AJM gratefully acknowledge support by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award #0000232253 and #DE-SC0019441, as part of the Computational Chemical Sciences Program. The work at Argonne National Laboratory was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences and Biosciences Division, under Contract No. DE-AC02-06CH11357, through the Computational Chemical Sciences Program (DHB). BK acknowledges financial support from the Alexander von Humboldt Foundation. The authors thank Zachary Ulissi for helpful discussions.

Notes and references

1 C. Chizallet and P. Raybaud, *Catal. Sci. Technol.*, 2014, **4**, 2797–2813.
2 P. Sit and L. Zhang, *Heterogeneous Catalysts: Advanced Design, Characterization and Applications*, 2021, **2**, 405–418.
3 T. Van Mourik, M. Bühl and M.-P. Gaigeot, *Density functional theory across chemistry, physics and biology*, 2014.
4 A. S. Rosen, J. M. Notestein and R. Q. Snurr, *J. Comput. Chem.*, 2019, **40**, 1305–1318.
5 K. D. Vogiatzis, M. V. Polynski, J. K. Kirkland, J. Townsend, A. Hashemi, C. Liu and E. A. Pidko, *Chem. Rev.*, 2018, **119**, 2453–2523.
6 A. Ishikawa, *Sci. Rep.*, 2022, **12**, 11657.
7 J. K. Nørskov, F. Abild-Pedersen, F. Studt and T. Bligaard, *Proc. Natl. Acad. Sci.*, 2011, **108**, 937–943.
8 C. A. Gaggioli, S. J. Stoneburner, C. J. Cramer and L. Gagliardi, *ACS Catal.*, 2019, **9**, 8481–8502.
9 G. D. Wehinger, M. Ambrosetti, R. Cheula, Z.-B. Ding, M. Isoz, B. Kreitz, K. Kuhlmann, M. Kutscherauer, K. Niyogi,

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

J. Poissonnier, R. Réocreux, D. Rudolf, J. Wagner, R. Zimmermann, M. Bracconi, H. Freund, U. Krewer and M. Maestri, *Chem. Eng. Res. Des.*, 2022, **184**, 39–58.

10 A. Bruix, J. T. Margraf, M. Andersen and K. Reuter, *Nat. Catal.*, 2019, **2**, 659–670.

11 J. Resasco, F. Abild-Pedersen, C. Hahn, Z. Bao, M. T. M. Koper and T. F. Jaramillo, *Nat. Catal.*, 2022, **5**, 374–381.

12 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder *et al.*, *APL Mater.*, 2013, **1**, 011002.

13 C. L. Zitnick, L. Chanussot, A. Das, S. Goyal, J. Heras-Domingo, C. Ho, W. Hu, T. Lavril, A. Palizhati, M. Riviere *et al.*, *arXiv preprint arXiv:2010.09435*, 2020, 1–27.

14 L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick and Z. Ulissi, *ACS Catal.*, 2021, **11**, 6059–6072.

15 R. Tran, J. Lan, M. Shuaibi, B. M. Wood, S. Goyal, A. Das, J. Heras-Domingo, A. Kolluru, A. Rizvi, N. Shoghi, A. Sriram, F. Therrien, J. Abed, O. Voznyy, E. H. Sargent, Z. Ulissi and C. L. Zitnick, *ACS Catal.*, 2023, **13**, 3066–3084.

16 C. Draxl and M. Scheffler, *MRS Bulletin*, 2018, **43**, 676–682.

17 C. Draxl and M. Scheffler, *J. Phys.: Mater.*, 2019, **2**, 036001.

18 M. Scheffler, M. Aeschlimann, M. Albrecht, T. Bereau, H.-J. Bungartz, C. Felser, M. Greiner, A. Groß, C. T. Koch, K. Kremer *et al.*, *Nature*, 2022, **604**, 635–642.

19 K. T. Winther, M. J. Hoffmann, J. R. Boes, O. Mamun, M. Bajdich and T. Bligaard, *Sci. Data*, 2019, **6**, 75.

20 S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. Hart, S. Sanvito, M. Buongiorno-Nardelli *et al.*, *Comput. Mater. Sci.*, 2012, **58**, 227–235.

21 J. S. Hummelshøj, F. Abild-Pedersen, F. Studt, T. Bligaard and J. K. Nørskov, *Angew. Chem. Int. Ed.*, 2012, **51**, 272–274.

22 J. E. Saal, S. Kirklin, M. Aykol, B. Meredig and C. Wolverton, *JOM*, 2013, **65**, 1501–1509.

23 S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl and C. Wolverton, *npj Comput. Mater.*, 2015, **1**, 1–15.

24 V. I. Hegde, C. K. H. Borg, Z. del Rosario, Y. Kim, M. Hutchinson, E. Antono, J. Ling, P. Saxe, J. E. Saal and B. Meredig, *Phys. Rev. Mater.*, 2023, **7**, 053805.

25 K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso *et al.*, *Science*, 2016, **351**, aad3000.

26 S. F. Sousa, P. A. Fernandes and M. J. Ramos, *J. Phys. Chem. A*, 2007, **111**, 10439–10452.

27 G. Kresse and J. Furthmüller, *Comput. Mater. Sci.*, 1996, **6**, 15–50.

28 G. Kresse and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169.

29 P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo *et al.*, *J. Phys.: Condens. Matter*, 2009, **21**, 395502.

30 P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni *et al.*, *J. Phys.: Condens. Matter*, 2017, **29**, 465901.

31 E. Aprà, E. J. Bylaska, W. A. De Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, H. J. J. Van Dam, Y. Alexeev, J. Anchell, V. Anisimov, F. W. Aquino, R. Atta-Fynn, J. Autschbach, N. P. Bauman, J. C. Becca, D. E. Bernholdt, K. Bhaskaran-Nair, S. Bogatko, P. Borowski, J. Boschen, J. Brabec, A. Bruner, E. Cauët, Y. Chen, G. N. Chuev, C. J. Cramer, J. Daily, M. J. O. Deegan, T. H. Dunning, M. Dupuis, K. G. Dyall, G. I. Fann, S. A. Fischer, A. Fonari, H. Früchtl, L. Gagliardi, J. Garza, N. Gawande, S. Ghosh, K. Glaesemann, A. W. Götz, J. Hammond, V. Helms, E. D. Hermes, K. Hirao, S. Hirata, M. Jacquelin, L. Jensen, B. G. Johnson, H. Jónsson, R. A. Kendall, M. Klemm, R. Kobayashi, V. Konkov, S. Krishnamoorthy, M. Krishnan, Z. Lin, R. D. Lins, R. J. Littlefield, A. J. Logsdail, K. Lopata, W. Ma, A. V. Marenich, J. Martin Del Campo, D. Mejia-Rodriguez, J. E. Moore, J. M. Mullin, T. Nakajima, D. R. Nascimento, J. A. Nichols, P. J. Nichols, J. Nieplocha, A. Otero-de-la-Roza, B. Palmer, A. Panyala, T. Pirojsirikul, B. Peng, R. Peverati, J. Pittner, L. Pollack, R. M. Richard, P. Sadayappan, G. C. Schatz, W. A. Shelton, D. W. Silverstein, D. M. A. Smith, T. A. Soares, D. Song, M. Swart, H. L. Taylor, G. S. Thomas, V. Tipparaju, D. G. Truhlar, K. Tsemekhman, T. Van Voorhis, Á. Vázquez-Mayagoitia, P. Verma, O. Villa, A. Vishnu, K. D. Vogiatzis, D. Wang, J. H. Weare, M. J. Williamson, T. L. Windus, K. Woliński, A. T. Wong, Q. Wu, C. Yang, Q. Yu, M. Zacharias, Z. Zhang, Y. Zhao and R. J. Harrison, *J. Chem. Phys.*, 2020, **152**, 184102.

32 J. J. Mortensen, A. H. Larsen, M. Kuisma, A. V. Ivanov, A. Taghizadeh, A. Peterson, A. Haldar, A. O. Dohn, C. Schäfer, E. Ö. Jónsson, E. D. Hermes, F. A. Nilsson, G. Kastlunger, G. Levi, H. Jónsson, H. Häkkinen, J. Fojt, J. Kangsabanik, J. Sødequist, J. Lehtomäki, J. Heske, J. Enkovaara, K. T. Winther, M. Dulak, M. M. Melander, M. Ovesen, M. Louhivuori, M. Walter, M. Gjerding, O. Lopez-Acevedo, P. Erhart, R. Warmbier, R. Würdemann, S. Kaappa, S. Latini, T. M. Boland, T. Bligaard, T. Skovhus, T. Susi, T. Maxson, T. Rossi, X. Chen, Y. L. A. Schmerwitz, J. Schiøtz, T. Olsen, K. W. Jacobsen and K. S. Thygesen, *J. Chem. Phys.*, 2024, **160**, 092503.

33 A. D. Bochevarov, E. Harder, T. F. Hughes, J. R. Greenwood, D. A. Braden, D. M. Philipp, D. Rinaldo, M. D. Halls, J. Zhang and R. A. Friesner, *Int. J. Quantum Chem.*, 2013, **113**, 2110–2142.

34 Q. Xu, A. Sharma, B. Comer, H. Huang, E. Chow, A. J. Medford, J. E. Pask and P. Suryanarayana, *SoftwareX*, 2021, **15**, 100709.

35 A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka,

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *J. Phys.: Condens. Matter*, 2017, **29**, 273002.

36 S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, *Comput. Mater. Sci.*, 2013, **68**, 314–319.

37 M. Capdevila-Cortada, Z. Łodziana and N. López, *ACS Catal.*, 2016, **6**, 8370–8379.

38 B. Meredig, A. Thompson, H. A. Hansen, C. Wolverton and A. van de Walle, *Phys. Rev. B*, 2010, **82**, 195128.

39 R. Sundararaman, D. Vigil-Fowler and K. Schwarz, *Chem. Rev.*, 2022, **122**, 10651–10674.

40 X.-H. Yang, Y.-B. Zhuang, J.-X. Zhu, J.-B. Le and J. Cheng, *WIREs Comput. Mol. Sci.*, 2022, **12**, e1559.

41 R. Sundararaman, W. A. Goddard and T. A. Arias, *J. Chem. Phys.*, 2017, **146**, 114104.

42 C. R. Tezak, N. R. Singstock, A. W. Alherz, D. Vigil-Fowler, C. A. Sutton, R. Sundararaman and C. B. Musgrave, *ACS Catal.*, 2023, **13**, 12894–12903.

43 S. M. R. Islam, F. Khezeli, S. Ringe and C. Plaisance, *J. Chem. Phys.*, 2023, **159**, 234117.

44 X. Andrade, J. Alberdi-Rodriguez, D. A. Strubbe, M. J. Oliveira, F. Nogueira, A. Castro, J. Muguerza, A. Arruabarrena, S. G. Louie, A. Aspuru-Guzik *et al.*, *J. Phys.: Condens. Matter*, 2012, **24**, 233202.

45 X. Gao, S. Bai, D. Fazzi, T. Niehaus, M. Barbatti and W. Thiel, *J. Chem. Theory. Comput.*, 2017, **13**, 515–524.

46 G. R. Iyer and B. M. Rubenstein, *J. Phys. Chem. A*, 2022, **126**, 4636–4646.

47 H.-Z. Ye and T. C. Berkelbach, *Ab Initio Surface Chemistry with Chemical Accuracy*, 2024.

48 C. Sheldon, J. Paier and J. Sauer, *J. Chem. Phys.*, 2021, **155**, 174702.

49 H. Kim, N.-K. Yu, N. Tian and A. J. Medford, *Assessing exchange-correlation functionals for heterogeneous catalysis of nitrogen species*, 2024, `https://arxiv.org/abs/2403.14482`.

50 B. R. Goldsmith, J. Esterhuizen, J.-X. Liu, C. J. Bartel and C. Sutton, *AIChe J.*, 2018, **64**, 2311–2323.

51 A. J. Medford, M. R. Kunz, S. M. Ewing, T. Borders and R. Fushimi, *ACS Catal.*, 2018, **8**, 7403–7429.

52 K. Suzuki, T. Toyao, Z. Maeno, S. Takakusagi, K.-i. Shimizu and I. Takigawa, *ChemCatChem*, 2019, **11**, 4537–4547.

53 M. Zhong, K. Tran, Y. Min, C. Wang, Z. Wang, C.-T. Dinh, P. De Luna, Z. Yu, A. S. Rasouli, P. Brodersen, S. Sun, O. Voznyy, C.-S. Tan, M. Askerka, F. Che, M. Liu, A. Seifitokaldani, Y. Pang, S.-C. Lo, A. Ip, Z. Ulissi and E. H. Sargent, *Nature*, 2020, **581**, 178–183.

54 T. Williams, K. McCullough and J. A. Lauterbach, *Chem. Mater.*, 2019, **32**, 157–165.

55 D. Roy, S. C. Mandal and B. Pathak, *ACS Appl. Mater. Interfaces*, 2021, **13**, 56151–56163.

56 S. Pablo-García, S. Morandi, R. A. Vargas-Hernández, K. Jorner, Ž. Ivković, N. López and A. Aspuru-Guzik, *Nat. Comput. Sci.*, 2023, **3**, 433–442.

57 P. Schlexer Lamoureux, K. T. Winther, J. A. Garrido Torres, V. Streibel, M. Zhao, M. Bajdich, F. Abild-Pedersen and T. Bligaard, *ChemCatChem*, 2019, **11**, 3581–3601.

58 V. Fung, J. Zhang, E. Juarez and B. G. Sumpter, *npj Comput. Mater.*, 2021, **7**, 84.

59 J. T. Margraf, H. Jung, C. Scheurer and K. Reuter, *Nat. Catal.*, 2023, **6**, 112–121.

60 R. Urrego-Ortiz, S. Builes and F. Calle-Vallejo, *ACS Catal.*, 2022, **12**, 4784–4791.

61 C. Carbogno, K. S. Thygesen, B. Bieniek, C. Draxl, L. M. Ghiringhelli, A. Gulans, O. T. Hofmann, K. W. Jacobsen, S. Lubeck, J. J. Mortensen *et al.*, *npj Comput. Mater.*, 2022, **8**, 69.

62 A. H. Motagamwala and J. A. Dumesic, *Chem. Rev.*, 2020, **121**, 1049–1076.

63 W. Xie, J. Xu, J. Chen, H. Wang and P. Hu, *Acc. Chem. Res.*, 2022, **55**, 1237–1248.

64 A. A. Gokhale, S. Kandoi, J. P. Greeley, M. Mavrikakis and J. A. Dumesic, *Chem. Eng. Sci.*, 2004, **59**, 4679–4691.

65 S. Bhandari, S. Rangarajan and M. Mavrikakis, *Acc. Chem. Res.*, 2020, **53**, 1893–1904.

66 T. Avanesian, S. Dai, M. J. Kale, G. W. Graham, X. Pan and P. Christopher, *J. Am. Chem. Soc.*, 2017, **139**, 4551–4558.

67 J. Wellendorff, T. L. Silbaugh, D. Garcia-Pintos, J. K. Nørskov, T. Bligaard, F. Studt and C. T. Campbell, *Surf. Sci.*, 2015, **640**, 36–44.

68 B. Kreitz, P. Lott, F. Studt, A. J. Medford, O. Deutschmann and C. F. Goldsmith, *Angew. Chem. Int. Ed.*, 2023, **62**, e202306514.

69 A. J. Medford, J. Wellendorff, A. Vojvodic, F. Studt, F. Abild-Pedersen, K. W. Jacobsen, T. Bligaard and J. K. Nørskov, *Science*, 2014, **345**, 197–200.

70 B. Kreitz, K. Sargsyan, K. Blöndal, E. J. Mazeau, R. H. West, G. D. Wehinger, T. Turek and C. F. Goldsmith, *JACS Au*, 2021, **1**, 1656–1673.

71 B. Kreitz, G. D. Wehinger, C. F. Goldsmith and T. Turek, *ChemCatChem*, 2022, **14**, e202200570.

72 J. E. Sutton, W. Guo, M. A. Katsoulakis and D. G. Vlachos, *Nat. Chem.*, 2016, **8**, 331–337.

73 P.J. Linstrom and W.G. Mallard, Eds., *NIST Chemistry WebBook, NIST Standard Reference Database Number 69, National Institute of Standards and Technology*, `https://doi.org/10.18434/T4D303`, (accessed 2023-11-30).

74 M. Chase, N. I. of Standards and T. (U.S.), *NIST-JANAF Thermochemical Tables*, American Chemical Society, 1998.

75 B. Ruscic and D. H. Bross, *Active Thermochemical Tables (ATcT) Values Based on ver. 1.130 of the Thermochemical Network*, `https://atct.anl.gov`, (accessed 2023-11-30).

76 B. Ruscic, *J. Phys. Chem. A*, 2015, **119**, 7810–7837.

77 L. P. Granda-Marulanda, A. Rendon-Calle, S. Builes, F. Illas, M. T. Koper and F. Calle-Vallejo, *ACS Catal.*, 2020, **10**, 6900–6907.

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

78 B. W. Chen, L. Xu and M. Mavrikakis, *Chem. Rev.*, 2020, **121**, 1007–1048.

79 C. P. Marshall, J. Schumann and A. Trunschke, *Angew. Chem. Int. Ed.*, 2023, **62**, e202302971.

80 C. R. Second, E. Edition, G. Ertl, H. Knözinger, F. Schüth, J. Weitkamp and W.-V. V. G. C. KGaA, *Handbook of heterogeneous catalysis*, 2008.

81 K. F. Kalz, R. Kraehnert, M. Dvoyashkin, R. Dittmeyer, R. Gläser, U. Krewer, K. Reuter and J.-D. Grunwaldt, *ChemCatChem*, 2017, **9**, 17–29.

82 M. Zhang, M. Wang, B. Xu and D. Ma, *Joule*, 2019, **3**, 2876–2883.

83 J. K. Nørskov, F. Studt, F. Abild-Pedersen and T. Bligaard, *Fundamental concepts in heterogeneous catalysis*, John Wiley & Sons, 2014.

84 A. V. Marenich, J. Ho, M. L. Coote, C. J. Cramer and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2014, **16**, 15068–15106.

85 J. D. Cox, *Pure Appl. Chem.*, 1982, **54**, 1239–1250.

86 B. Ruscic and D. H. Bross, *Mathematical Modelling of Gas-Phase Complex Reaction Systems: Pyrolysis and Combustion*, Elsevier, 2019, vol. 45, pp. 3–114.

87 A. Karton, E. Rabinovich, J. M. L. Martin and B. Ruscic, *J. Chem. Phys.*, 2006, **125**, 144108.

88 A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez and J. F. Stanton, *J. Chem. Phys.*, 2004, **121**, 11599–11613.

89 H. M. Jaeger, H. F. Schaefer, J. Demaison, A. G. Császár and W. D. Allen, *J. Chem. Theory Comput.*, 2010, **6**, 3066–3078.

90 S. J. Klippenstein, L. B. Harding and B. Ruscic, *J. Phys. Chem. A*, 2017, **121**, 6580–6602.

91 B. Ruscic, R. E. Pinzon, G. Von Laszewski, D. Kodeboyina, A. Burcat, D. Leahy, D. Montoy and A. F. Wagner, Journal of Physics: Conference Series, 2005, p. 561.

92 B. Ruscic, R. E. Pinzon, M. L. Morton, G. von Laszevski, S. J. Bittner, S. G. Nijsure, K. A. Amin, M. Minkoff and A. F. Wagner, *J. Phys. Chem. A*, 2004, **108**, 9979–9997.

93 K. Reuter, *Catal. Lett.*, 2016, **146**, 541–563.

94 K. Reuter and M. Scheffler, *Phys. Rev. B*, 2001, **65**, 035406.

95 T. L. Silbaugh and C. T. Campbell, *J. Phys. Chem. C*, 2016, **120**, 25161–25172.

96 R. B. Araujo, G. L. S. Rodrigues, E. C. dos Santos and L. G. M. Pettersson, *Nat. Commun.*, 2022, **13**, 6853.

97 R. S. Kingsbury, A. S. Rosen, A. S. Gupta, J. M. Munro, S. P. Ong, A. Jain, S. Dwaraknath, M. K. Horton and K. A. Persson, *npj Comput. Mater.*, 2022, **8**, 195.

98 B. T. G. Lau, G. Knizia and T. C. Berkelbach, *J. Phys. Chem. Lett.*, 2021, **12**, 1104–1109.

99 F. Libisch, C. Huang and E. A. Carter, *Acc. Chem. Res.*, 2014, **47**, 2768–2775.

100 Q. Zhao and E. A. Carter, *J. Chem. Theory Comput.*, 2020, **16**, 6528–6538.

101 L. Schimka, J. Harl, A. Stroppa, A. Grüneis, M. Marsman, F. Mittendorfer and G. Kresse, *Nat. Mater.*, 2010, **9**, 741–744.

102 N. A. Szaro, M. Bello, C. H. Fricke, O. H. Bamidele and A. Heyden, *J. Phys. Chem. Lett.*, 2023, **14**, 10769–10778.

103 X. Ren, P. Rinke and M. Scheffler, *Phys. Rev. B*, 2009, **80**, 045402.

104 P. S. Schmidt and K. S. Thygesen, *J. Phys. Chem. C*, 2018, **122**, 4381–4390.

105 J. A. Garrido Torres, B. Ramberger, H. A. Früchtl, R. Schaub and G. Kresse, *Phys. Rev. Mater.*, 2017, **1**, 060803.

106 P. N. Plessow, A. Smith, S. Tischer and F. Studt, *J. Am. Chem. Soc.*, 2019, **141**, 5908–5915.

107 P. N. Plessow, A. E. Enss, P. Huber and F. Studt, *Catal. Sci. Technol.*, 2022, **12**, 3516–3523.

108 S. Svelle, C. Tuma, X. Rozanska, T. Kerber and J. Sauer, *J. Am. Chem. Soc.*, 2009, **131**, 816–825.

109 N. Hansen, T. Kerber, J. Sauer, A. T. Bell and F. J. Keil, *J. Am. Chem. Soc.*, 2010, **132**, 11525–11538.

110 R. O. Sharma, T. T. Rantala and P. E. Hoggan, *J. Phys. Chem. C*, 2020, **124**, 26232–26240.

111 J. Sauer, *Acc. Chem. Res.*, 2019, **52**, 3502–3510.

112 C. T. Campbell and J. R. V. Sellers, *J. Am. Chem. Soc.*, 2012, **134**, 18109–18115.

113 G. Collinge, S. F. Yuk, M.-T. Nguyen, M.-S. Lee, V.-A. Glezakou and R. Rousseau, *ACS Catal.*, 2020, **10**, 9236–9260.

114 L. H. Sprowl, C. T. Campbell and L. Árnadóttir, *J. Phys. Chem. C*, 2016, **120**, 9719–9731.

115 K. Blöndal, K. Sargsyan, D. H. Bross, B. Ruscic and C. F. Goldsmith, *J. Phys. Chem. C*, 2021, **125**, 20249–20260.

116 K. Blöndal, K. Sargsyan, D. H. Bross, B. Ruscic and C. F. Goldsmith, *ACS Catal.*, 2023-01-06, **13**, 19–32.

117 C. Waitt, A. R. Miles and W. F. Schneider, *J. Phys. Chemi. C*, 2021, **125**, 20331–20342.

118 C. T. Campbell, *Acc. Chem. Res.*, 2019, **52**, 984–993.

119 S. L. Tait, Z. Dohnálek, C. T. Campbell and B. D. Kay, *J. Chem. Phys.*, 2006, **125**, 234308.

120 E. M. Karp, T. L. Silbaugh and C. T. Campbell, *J. Phys. Chem. C*, 2013, **117**, 6325–6336.

121 E. M. Karp, T. L. Silbaugh and C. T. Campbell, *J. Am. Chem. Soc.*, 2013, **135**, 5208–5211.

122 E. M. Karp, T. L. Silbaugh, M. C. Crowe and C. T. Campbell, *J. Am. Chem. Soc.*, 2012, **134**, 20388–20395.

123 A. B. Mhadeshwar, H. Wang and D. G. Vlachos, *J. Phys. Chem. B*, 2003, **107**, 12721–12733.

124 B. Kreitz, G. S. Gusmão, D. Nai, S. J. Sahoo, A. A. Peterson, D. H. Bross, C. F. Goldsmith and A. J. Medford, *Data for Unifying thermochemistry concepts in computational heterogeneous catalysis*, 2024, `https://doi.org/10.5281/zenodo.13143954`.

125 M. D. Porosoff, M. N. Z. Myint, S. Kattel, Z. Xie, E. Gomez, P. Liu and J. G. Chen, *Angew. Chem. Int. Ed.*, 2015, **54**, 15501–15505.

126 M. Myint, B. Yan, J. Wan, S. Zhao and J. G. Chen, *J. Catal.*, 2016, **343**, 168–177.

127 F. Jalid, T. S. Khan and M. A. Haider, *Catal. Sci. Technol.*, 2021, **11**, 97–115.

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

128 B. Kreitz, K. Abeywardane and C. F. Goldsmith, *J. Chem. Theory Comput.*, 2023, **19**, 4149–4162.

129 J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard and K. W. Jacobsen, *Phys. Rev. B*, 2012, **85**, 235149.

130 B. Hammer, L. B. Hansen and J. K. Nørskov, *Phys. Rev. B*, 1999, **59**, 7413–7421.

131 D. W. Flaherty, D. D. Hibbitts, E. I. Gürbüz and E. Iglesia, *J. Catal.*, 2014, **311**, 350–356.

132 A. Almithn and D. Hibbitts, *ACS Catal.*, 2018, **8**, 6375–6387.

133 D. D. Hibbitts, D. W. Flaherty and E. Iglesia, *ACS Catal.*, 2016, **6**, 469–482.

134 Z. W. Ulissi, A. J. Medford, T. Bligaard and J. K. Nørskov, *Nat. Commun.*, 2017, **8**, 14621.

135 H. Li and K. Reuter, *ACS Catal.*, 2022, **12**, 10506–10513.

136 T. P. Senftle, A. C. T. Van Duin and M. J. Janik, *ACS Catal.*, 2015, **5**, 6187–6199.

137 D. Opalka, C. Scheurer and K. Reuter, *ACS Catal.*, 2019, **9**, 4944–4950.

138 N. Bergmann, N. G. Hörmann and K. Reuter, *J. Chem. Theory Comput.*, 2023, **19**, 8815–8825.

139 A. J. Medford, A. Vojvodic, F. Studt, F. Abild-Pedersen and J. K. Nørskov, *J. Catal.*, 2012, **290**, 108–117.

140 A. Banerjee, V. Navarro, J. W. M. Frenken, A. P. van Bavel, H. P. C. E. Kuipers and M. Saeys, *J. Phys. Chem. Lett.*, 2016, **7**, 1996–2001.

141 S. P. Vicchio, Z. Chen, K. W. Chapman and R. B. Getman, *J. Am. Chem. Soc.*, 2023, **145**, 2852–2859.

142 R. Cheula, A. Soon and M. Maestri, *Catal. Sci. Technol.*, 2018, **8**, 3493–3503.

143 J. K. Nørskov, J. Rossmeisl, A. Logadottir, L. Lindqvist, J. R. Kitchin, T. Bligaard and H. Jónsson, *J. Phys. Chem. B*, 2004, **108**, 17886–17892.

144 A. A. Peterson, F. Abild-Pedersen, F. Studt, J. Rossmeisl and J. K. Nørskov, *Energy Environ. Sci.*, 2010, **3**, 1311.

145 P. Lindgren, G. Kastlunger and A. A. Peterson, *J. Chem. Phys.*, 2022, **157**, 180902.

146 A. J. Medford, A. C. Lausche, F. Abild-Pedersen, B. Temel, N. C. Schjødt, J. K. Nørskov and F. Studt, *Top. Catal.*, 2014, **57**, 135–142.

147 A. C. Lausche, A. J. Medford, T. S. Khan, Y. Xu, T. Bligaard, F. Abild-Pedersen, J. K. Nørskov and F. Studt, *J. Catal.*, 2013, **307**, 275–282.

148 S. Wang, V. Petzold, V. Tripkovic, J. Kleis, J. G. Howalt, E. Skúlason, E. M. Fernández, B. Hvolbæk, G. Jones, A. Toftelund, H. Falsig, M. Björketun, F. Studt, F. Abild-Pedersen, J. Rossmeisl, J. K. Nørskov and T. Bligaard, *Phys. Chem. Chem. Phys.*, 2011, **13**, 20760.

149 M. Andersen, C. P. Plaisance and K. Reuter, *J. Chem. Phys.*, 2017, **147**, 152705.

150 N. Yang, A. J. Medford, X. Liu, F. Studt, T. Bligaard, S. F. Bent and J. K. Nørskov, *J. Am. Chem. Soc.*, 2016, **138**, 3705–3714.

151 J. L. Snider, V. Streibel, M. A. Hubert, T. S. Choksi, E. Valle, D. C. Upham, J. Schumann, M. S. Duyar, A. Gallo, F. Abild-Pedersen and T. F. Jaramillo, *ACS Catal.*, 2019, **9**, 3399–3412.

152 G. T. K. K. Gunasooriya, A. P. van Bavel, H. P. C. E. Kuipers and M. Saeys, *ACS Catal.*, 2016, **6**, 3660–3664.

153 R. García-Muelas and N. López, *Nat. Commun.*, 2019, **10**, 4687.

154 C. A. Wolcott, A. J. Medford, F. Studt and C. T. Campbell, *J. Catal.*, 2015, **330**, 197–207.

155 M. H. Hansen, J. K. Nørskov and T. Bligaard, *J. Catal.*, 2019, **374**, 161–170.

156 A. A. Latimer, A. R. Kulkarni, H. Aljama, J. H. Montoya, J. S. Yoo, C. Tsai, F. Abild-Pedersen, F. Studt and J. K. Nørskov, *Nat. Mater.*, 2017, **16**, 225–229.

157 C. Fricke, B. Rajbanshi, E. A. Walker, G. Terejanu and A. Heyden, *ACS Catal.*, 2022, **12**, 2487–2498.

158 G. Kastlunger, L. Wang, N. Govindarajan, H. H. Heenen, S. Ringe, T. Jaramillo, C. Hahn and K. Chan, *ACS Catal.*, 2022, **12**, 4344–4357.

159 M. Valter, E. C. Dos Santos, L. G. M. Pettersson and A. Hellman, *J. Phys. Chem. C*, 2020, **124**, 17907–17915.

160 A. J. Medford, C. Shi, M. J. Hoffmann, A. C. Lausche, S. R. Fitzgibbon, T. Bligaard and J. K. Nørskov, *Catal. Lett.*, 2015, **145**, 794–807.

161 H. Wu, A. M. Payne, H.-W. Pang, A. Menon, C. A. Grambow, D. S. Ranasinghe, X. Dong, A. Grinberg Dana and W. H. Green, *J. Phys. Chem. A*, 2024, **128**, 4335–4352.

162 A. Jain, G. Hautier, S. P. Ong, C. J. Moore, C. C. Fischer, K. A. Persson and G. Ceder, *Phys. Rev. B*, 2011, **84**, 045115.

163 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. Von Lilienfeld, *J. Chem. Theory. Comput.*, 2015, **11**, 2087–2096.

164 H. Bhattacharjee and D. G. Vlachos, *J. Chem. Inf. Model.*, 2020, **60**, 4673–4683.

165 H. Bhattacharjee, N. Anesiadis and D. G. Vlachos, *Sci. Rep.*, 2021, **11**, 14372.

166 A. J. Cohen, P. Mori-Sánchez and W. Yang, *Chem. Rev.*, 2012, **112**, 289–320.

167 C. F. Goldsmith, G. R. Magoon and W. H. Green, *J. Phys. Chem. A*, 2012, **116**, 9033–9057.

168 R. Urrego-Ortiz, S. Builes and F. Calle-Vallejo, *ChemCatChem*, 2021, **13**, 2508–2516.

169 C. P. Plaisance and R. A. Van Santen, *J. Am. Chem. Soc.*, 2015, **137**, 14660–14672.

170 E. Sargeant, F. Illas, P. Rodríguez and F. Calle-Vallejo, *Electrochi. Acta*, 2022, **426**, 140799.

171 F. Studt, F. Abild-Pedersen, J. B. Varley and J. K. Nørskov, *Catal. Lett.*, 2013, **143**, 71–73.

172 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 and R. Schlögl, *ChemCatChem*, 2015, **7**, 1105–1111.

173 V. Streibel, H. A. Aljama, A.-C. Yang, T. S. Choksi, R. S. Sánchez-Carrera, A. Schäfer, Y. Li, M. Cargnello and F. Abild-Pedersen, *ACS Catal.*, 2022, **12**, 1742–1757.

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

174 B. Zijlstra, R. J. Broos, W. Chen, I. A. Filot and E. J. Hensen, *Catal. Today*, 2020, **342**, 131–141.

175 L. C. Grabow and M. Mavrikakis, *ACS Catal.*, 2011, **1**, 365–384.

176 R. Cheula and M. Maestri, *Catal. Today*, 2021, S0920586121002261.

177 R. Urrego-Ortiz, S. Builes, F. Illas and F. Calle-Vallejo, *EES Catal.*, 2024, **2**, 157–179.

178 C. A. Grambow, Y.-P. Li and W. H. Green, *J. Phys. Chem. A*, 2019, **123**, 5826–5835.

179 R. Christensen, H. A. Hansen and T. Vegge, *Catal. Sci. Technol.*, 2015, **5**, 4946–4949.

180 G. H. Gu, P. Plechac and D. G. Vlachos, *React. Chem. Eng.*, 2018, **3**, 454–466.

181 K. Blöndal, J. Jelic, E. Mazeau, F. Studt, R. H. West and C. F. Goldsmith, *Ind. Eng. Chem. Res.*, 2019, **58**, 17682–17691.

182 S. E. Wheeler, K. N. Houk, P. v. R. Schleyer and W. D. Allen, *J. Am. Chem. Soc.*, 2009, **131**, 2547–2560.

183 V. Vorotnikov, S. Wang and D. G. Vlachos, *Ind. Eng. Chem. Res.*, 2014, **53**, 11929–11938.

184 V. Vorotnikov and D. G. Vlachos, *J. Phys. Chem. C*, 2015, **119**, 10417–10426.

185 G. H. Gu, B. Schweitzer, C. Michel, S. N. Steinmann, P. Sautet and D. G. Vlachos, *J. Phys. Chem. C*, 2017, **121**, 21510–21519.

186 G. H. Gu and D. G. Vlachos, *J. Phys. Chem. C*, 2016, **120**, 19234–19241.

187 B. Kreitz, P. Lott, J. Bae, K. Blöndal, S. Angeli, Z. W. Ulissi, F. Studt, C. F. Goldsmith and O. Deutschmann, *ACS Catal.*, 2022, **12**, 11137–11151.

188 B. Kreitz, K. Blöndal, K. Badger, R. H. West and C. F. Goldsmith, *Digital Discovery*, 2024, **3**, 173–185.

189 E. J. Mazeau, P. Satpute, K. Blöndal, C. F. Goldsmith and R. H. West, *ACS Catal.*, 2021, **11**, 7114–7125.

190 B. Kreitz, G. D. Wehinger, C. F. Goldsmith and T. Turek, *J. Phys. Chem. C*, 2021, **125**, 2984–3000.

191 D. W. Blaylock, T. Ogura, W. H. Green and G. J. O. Beran, *J. Phys. Chem. C*, 2009, **113**, 4898–4908.

192 D. W. Blaylock, Y.-A. Zhu and W. H. Green, *Top. Catal.*, 2011, **54**, 828–844.

193 J. T. Stuckless, N. Al-Sarraf, C. Wartnaby and D. A. King, *J. Chem. Phys.*, 1993, **99**, 2202–2212.

194 J. T. Stuckless, C. E. Wartnaby, N. Al-Sarraf, St. J. B. Dixon-Warren, M. Kovar and D. A. King, *J. Chem. Phys.*, 1997, **106**, 2012–2030.

195 J. Lapujoulade and K. S. Neil, *J. Chem. Phys.*, 1972, **57**, 3535–3545.

196 J. Segner, C. Campbell, G. Doyen and G. Ertl, *Surf. Sci.*, 1984, **138**, 505–523.

197 W. J. Hehre, R. Ditchfield, L. Radom and J. A. Pople, *J. Am. Chem. Soc.*, 1970, **92**, 4796–4801.

198 J. A. Pople, L. Radom and W. J. Hehre, *J. Am. Chem. Soc*, 1971, **93**, 289–300.

199 R. O. Ramabhadran and K. Raghavachari, *J. Chem. Theory Comput.*, 2011, **7**, 2094–2103.

200 R. O. Ramabhadran and K. Raghavachari, *Acc. Chem. Res.*, 2014, **47**, 3596–3604.

201 S. E. Wheeler, *WIREs Comput. Mol. Sci.*, 2012, **2**, 204–220.

202 S. N. Elliott, M. Keçeli, M. K. Ghosh, K. P. Somers, H. J. Curran and S. J. Klippenstein, *J. Phys. Chem. A*, 2023, acs.jpca.2c07248.

203 P. Buerger, J. Akroyd, S. Mosbach and M. Kraft, *Combust. Flame*, 2018, **187**, 105–121.

204 J. Kubota, S. Ichihara, J. N. Kondo, K. Domen and C. Hirose, *Surf. Sci.*, 1996, 5.

205 N. M. Abbas and R. J. Madix, *Appl. Surf. Sci.*, 1981, **7**, 241–275.

206 C. A. Wolcott, I. X. Green, T. L. Silbaugh, Y. Xu and C. T. Campbell, *J. Phys. Chem. C*, 2014, **118**, 29310–29321.

207 E. M. Karp, C. T. Campbell, F. Studt, F. Abild-Pedersen and J. K. Nørskov, *J. Phys. Chem. C*, 2012, **116**, 25772–25776.

208 K. Abeywardane and C. F. Goldsmith, *ACS Phys. Chem. Au*, 2024, **4**, 247–258.

209 E. H. Moore, *Bulletin of the American Mathematical Society*, 1920, **26**, 394–395.

210 A. Bjerhammar, *Bulletin géodésique*, 1951, **20**, 188–220.

211 R. Penrose, *Math. Proc. Cambridge Philos. Soc.*, 1955, **51**, 406–413.

https://doi.org/10.26434/chemrxiv-2024-fjfmd **ORCID:** https://orcid.org/0000-0003-0158-9147 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0