



On the Challenge of Obtaining an Accurate Solvation Energy Estimate in Simulations of Electrocatalysis

Björn Kirchhoff^{1,2} · Elvar Ö. Jónsson² · Timo Jacob^{1,3,4} · Hannes Jónsson²

Accepted: 5 May 2023
© The Author(s) 2023

Abstract

The effect of solvation on the free energy of reaction intermediates adsorbed on electrocatalyst surfaces can significantly change the thermochemical overpotential, but accurate calculations of this are challenging. Here, we present computational estimates of the solvation energy for reaction intermediates in oxygen reduction reaction (ORR) on a B-doped graphene (BG) model system where the overpotential is found to reduce by up to 0.6 V due to solvation. BG is experimentally reported to be an active ORR catalyst but recent computational estimates using state-of-the-art hybrid density functionals in the absence of solvation effects have indicated low activity. To test whether the inclusion of explicit solvation can bring the calculated activity estimates closer to the experimental reports, up to 4 layers of water molecules are included in the simulations reported here. The calculations are based on classical molecular dynamics and local minimization of energy using atomic forces evaluated from electron density functional theory. Data sets are obtained from regular and coarse-grained dynamics, as well as local minimization of structures resampled from dynamics simulations. The results differ greatly depending on the method used and the solvation energy estimates are deemed untrustworthy. It is concluded that a significantly larger number of water molecules is required to obtain converged results for the solvation energy. As the present system includes up to 139 atoms, it already strains the limits of computational feasibility, so this points to the need for a hybrid simulation approach where efficient simulations of much larger number of solvent molecules is carried out using a lower level of theory while retaining the higher level of theory for the reacting molecules as well as their near neighbors and the catalyst. The results reported here provide a word of caution to the computational catalysis community: activity predictions can be inaccurate if too few solvent molecules are included in the calculations.

Keywords Solvation · Electrochemistry · Oxygen reduction reaction · Doped graphene

1 Introduction

The replacement of costly and rare precious metals with cheaper and more abundant elements in catalysts, for example in the oxygen reduction reaction (ORR) in fuel cells, is an important milestone towards sustainable energy

production. To this end, heteroatom-doped graphenes have been explored extensively [1–3] following experiments showing high ORR activity of a nitrogen-doped graphene (NG) electrocatalyst in 2010 [4]. Soon after the reports of high catalytic activity of NG, boron-doped graphene (BG) emerged as another promising candidate for efficient ORR electrocatalysis.

Sheng et al. [5] measured favorable alkaline ORR activity for BG with 3.2% dopant concentration synthesized using Hummer's method [6, 7]. Their BG material catalyzed the $4e^-$ ORR pathway and showed good tolerance to CO poisoning. Note that Hummer's method has become subject to criticism as it can deposit significant amounts of transition metal impurities in the material [8, 9] which cannot be removed using typical wet-chemical purification methods [10]. In the same vein, Xu et al. [11] and Jiao et al. [12] synthesized NG and BG using Hummer's

✉ Björn Kirchhoff
bjoern.kirchhoff@uni-ulm.de

¹ Institute of Electrochemistry, Ulm University, Albert-Einstein-Allee 47, 89081 Ulm, Germany

² Science Institute and Faculty of Physical Sciences, University of Iceland, VR-III, Hjarðarhagi 2, 107 Reykjavík, Iceland

³ Helmholtz-Institute Ulm (HIU) Electrochemical Energy Storage, Helmholtz-Straße 11, 89081 Ulm, Germany

⁴ Karlsruhe Institute of Technology (KIT), P.O. Box 3640, 76021 Karlsruhe, Germany

method. Both groups report that NG and BG are efficient ORR catalysts, showing similarly high ORR activity in their experiments and corresponding calculations. Further experimental work is summarized in a 2016 review by Agnoli and Favaro [3].

Computational predictions of the ORR activity of BG have overall been promising. The free energy approach using the computational hydrogen electrode (CHE) [13] is often used to evaluate the ORR activity of computational models. Since the estimate of an overpotential obtained by this approach only reflects thermodynamic free energy of intermediates as well as initial and final states, it will be referred to as the thermochemical overpotential, η_{TCM} , in the following.

Jiao and co-workers predict a η_{TCM} range of 0.4–0.6 V for both BG and NG based on calculations using the B3LYP functional and molecular flake model systems, in good agreement with their experimental measurements [12]. A similar value, 0.38 V, is reported by Wang et al. [14] for a BG nanoribbon model using the PBE functional and DFT-D3 [15, 16] dispersion correction. The most optimistic prediction is reported by Fazio and co-workers with a η_{TCM} of 0.29 V in a B3LYP-based study of a BG flake model system [17]. For reference, the measured overpotential of a typical Pt/C electrocatalyst is 0.3–0.4 V [18]. The experimental overpotential, however, depends on many other factors besides adsorption strength of the ORR intermediates, hence η_{TCM} values are only a rough and purely thermodynamic approximation of the actual overpotential.

The exact mechanism of the ORR on BG is a matter of ongoing investigation. Fazio and co-workers established that the associative $4e^-$ pathway should be dominant for BG from a theoretical perspective [17]. They found O_2 adsorption to occur via an open-shell end-on intermediate on a molecular flake model system in calculations using the B3LYP functional. Ferrighi et al. proposed the formation of stable B– O_3 bulk oxides on BG which they hypothesize to be the first step in the ORR mechanism on BG [19]. They, however, did not detail further reaction steps. Ferrighi et al. used a molecular flake model and the B3LYP functional as well as periodic surface models and the PBE functional in their study. Contrarily, Wang and co-workers recently identified a cluster of two B dopants in *para* arrangement to enable the associative $4e^-$ ORR pathway, including energetically favorable O_2 adsorption [14]. They used a periodic nanoribbon model and the PBE functional with DFT-D3 dispersion correction. Using a molecular flake model and the B3LYP functional, the study by Jiao et al. [12] finds that a top adsorption geometry should be favored for the critical *O intermediate on BG while other studies [14, 17, 19] typically find a B–C bridge site to be favored for *O adsorption. It can be summarized that the active site debate for the ORR mechanism on BG is not settled yet.

Furthermore, the stabilization of the ORR intermediates on BG by water molecules, which has been found to be a significant contribution to the free energy description of the ORR on NG, [20–23] has only been considered by one group so far to the best of the authors' knowledge. Fazio et al. used a cluster of 6 water molecules in contact with a molecular flake model representing BG to estimate the effects of solvation [17]. The group found that while the stability of the *O intermediate is barely affected by solvation, the *OH and *OOH intermediates are stabilized by -0.37 eV and -0.46 eV, respectively. The low predicted η_{TCM} of 0.29 V versus SHE in this study results in part from the stabilizing effect of solvation.

In the study by Jiao et al. [12] solvation effects are estimated using implicit [24] solvation models. However, implicit solvation models have in some cases been shown to fail at reproducing experimental solvation energy measurements or solvation energy results from simulations using many explicit solvent molecules [25–28].

We recently presented results for the ORR on NG where it was shown that high-level DFT calculations based on hybrid functionals yield a η_{TCM} estimate close to 1.0 V versus SHE [29], which indicates catalytic inactivity. The choice of hybrid functional was made after benchmarking various functionals against a reference data set from diffusion Monte Carlo simulations [30].

However, it was noted that solvation effects could considerably improve the catalyst activity predictions. To illustrate this effect, we applied two sets of solvation stabilization energy, $\Delta\Delta E_{\text{solv}}$, data for the ORR intermediates on NG taken from literature sources (Reda et al. [23] and Yu et al. [21]) to the hybrid DFT free energy results. Solvation was found to reduce η_{TCM} by up to 0.5 V. However, the published $\Delta\Delta E_{\text{solv}}$ data set were calculated in different ways and disagreed significantly, leading to different η_{TCM} estimates depending on the choice of $\Delta\Delta E_{\text{solv}}$ data set.

The accurate hybrid DFT approach was also applied to BG with similar results: a η_{TCM} estimate above 1.0 V versus SHE, indicating catalytic inactivity [31]. This result is in stark contrast to other more optimistic studies which, importantly, used functionals such as PBE and B3LYP as well as molecular flake models which were shown to produce unreliable adsorption free energy results [29]. However, the high η_{TCM} prediction for BG did not include any solvation effects. Informed by the report from Fazio et al. on the significant impact of $\Delta\Delta E_{\text{solv}}$ on the free energy trends and by our own observations of the same for NG, the present study was conceived to systematically investigate the effect of an increasing number of explicit water molecules on the stability of the ORR intermediates *O, *OH, and *OOH, as represented by the $\Delta\Delta E_{\text{solv}}$ descriptor. Simulations were performed with the 32-atom BG model system used previously [31] in contact with up to 4 layers (32 molecules) of water.

Both local minimization calculations as well as regular and coarse-grained classical molecular dynamics (MD) simulations were performed using atomic forces derived from density functional theory (DFT) calculations to obtain statistical estimates of $\Delta\Delta E_{\text{solv}}$. Additionally, local minimization calculations were performed on structures re-sampled from the MD data sets. In short, none of the data sets generated in this way yielded converged and trustworthy $\Delta\Delta E_{\text{solv}}$ results. Technical aspects of the simulations are discussed in detail and the conclusion is that a much larger number of water molecules needs to be included in the calculations to provide reliable estimates of the solvation effect.

The present model system includes up to 139 atoms and the dynamics simulations span up to 100 ps, thereby already straining the computational resources. Moreover, the $\Delta\Delta E_{\text{solv}}$ estimates are highly system dependent and would need to be reestablished for every new (electro-) catalyst model. Hence, we highlight the need for hybrid simulation methods that enable simulations of systems including hundreds or even thousands of water molecules using a lower level of theory while retaining electronic structure level accuracy in the surface region where reactions occur.

2 Methodology

2.1 Calculation of the Solvation Stabilization Energy

The solvation stabilization energy $\Delta\Delta E_{\text{solv}}$ is estimated as the difference between the adsorption energy calculated for models in contact with explicit solvent ($\Delta E_{\text{ads}}^{\text{with solvent}}$) and models without inclusion of any solvent molecules ($\Delta E_{\text{ads}}^{\text{without solvent}}$):

$$\Delta\Delta E_{\text{solv}} = \Delta E_{\text{ads}}^{\text{with solvent}} - \Delta E_{\text{ads}}^{\text{without solvent}}, \quad (1)$$

where

$$\Delta E_{\text{ads}}^{\text{with solvent}} = E_{\text{tot}}^{\text{BG+adatom with solvent}} - E_{\text{tot}}^{\text{BG with solvent}} - E_{\text{tot}}^{\text{adatom reference}} \quad (2)$$

and

$$\Delta E_{\text{ads}}^{\text{without solvent}} = E_{\text{tot}}^{\text{BG+adatom without solvent}} - E_{\text{tot}}^{\text{BG without solvent}} - E_{\text{tot}}^{\text{adatom reference}}. \quad (3)$$

Here, $E_{\text{tot}}^{\text{adatom reference}}$ is the total energy of any combination of gasphase molecules used to calculate the adsorption energy. For example, $E_{\text{tot}}^{\text{adatom reference}}$ may be expanded to $E_{\text{tot}}^{\text{H}_2\text{O}} - E_{\text{tot}}^{\text{H}_2}$ to serve as the reference energy for an O adatom. Because these values are always gasphase reference energy values,

also in the case of the solvated model systems, they cancel out in the calculation of $\Delta\Delta E_{\text{solv}}$.

Therefore, Eq. (1) reduces to:

$$\Delta\Delta E_{\text{solv}} = E_{\text{tot}}^{\text{BG+adatom with solvent}} - E_{\text{tot}}^{\text{BG with solvent}} - (E_{\text{tot}}^{\text{BG+adatom without solvent}} - E_{\text{tot}}^{\text{BG without solvent}}) \quad (4)$$

2.2 Calculation of the Confidence Interval for Average Ensemble Properties

The confidence interval (CI) is a useful statistical measure for the error bar of an average result sampled from a normal distribution of values. It is therefore also useful to estimate the error bar of ensemble averages sampled through molecular dynamics integration; see Grossfield et al. [32] for more details. The CI defines an interval in which the true ensemble average lies with a certain probability. Here, a 95% probability threshold is used to define the error bars, i.e., the 95% CI.

The two-sided CI $\langle x \rangle$ of a variable x is defined as

$$\langle x \rangle = \bar{x} \pm U, \quad (5)$$

where \bar{x} is the ensemble average and U is the expanded uncertainty. The expanded uncertainty is defined as

$$U = k s(\bar{x}), \quad (6)$$

where k is the coverage factor and $s(\bar{x})$ is the experimental standard deviation of the mean. $s(\bar{x})$ is defined as

$$s(\bar{x}) = \frac{s(x)}{\sqrt{n}}, \quad (7)$$

where $s(x)$ is the experimental standard deviation

$$s(x) = \sqrt{\frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n - 1}} \quad (8)$$

with the sample values x_j , the arithmetic mean of the ensemble property \bar{x} , and the number of independent samples n (Table 1).

The coverage factor k is a measure for the number of independent samples taken into account during calculation of the standard deviation. For the 95% CI used in this work, the coverage factors k are given by Grossfield et al. as follows:

Table 1 Coverage factors k as a function of the number of independent samples n

| n | k |
|-----|------|
| 6 | 2.57 |
| 11 | 2.23 |
| 16 | 2.13 |
| 21 | 2.09 |
| 26 | 2.06 |
| 51 | 2.01 |
| 101 | 1.98 |

Reproduced from Grossfield et al. [32]

3 Computational Details

3.1 BG Sheet Model System

The model system used in this study is a 32-atomic graphene sheet with one B dopant atom, analogous to our previous works on NG and BG [29, 31]. To study the influence of solvation on the ORR intermediates $*O$, $*OH$, and $*OOH$, 1–4 layers of water molecules with 8 water molecules per layer are added to the model. The water configurations built initially were inspired by the configurations presented by Reda et al. in a study of the solvation of ORR intermediates on NG [23]. The group showed that the maximum H_2O coverage per layer for NG is $\Theta_{H_2O} = \frac{2}{3}$ monolayers which the present results confirm. Hence, a maximum of 24 atoms (8 molecules) can be placed per layer before lateral crowding destabilizes the water configuration and formation of a new layer begins. Figure 1 shows a representative illustration of the BG sheet model with an $*O$ adatom in contact with 4 layers of water molecules; illustrations of sheet models with $*OH$ and $*OOH$ ad molecules as well as models in contact with 1–3 layers of water are shown in Figs. S1 and S2, respectively.

In agreement with studies by Fazio et al. [17] Ferrighi et al. [19] and Wang et al. [14] but in disagreement with the study by Jiao et al. [12] we find adsorption of the $*O$ intermediate on the C–B bridge position to be energetically most favorable. The $*OH$ and $*OOH$ adspecies are found to adsorb most favorably on the B top position, which is in agreement with all previously mentioned studies.

The 32-atomic BG model system is converged with respect to the adsorption energy of the ORR intermediates, see Fig. S4. This model therefore allows for the study of the adsorption energy—and the influence of solvation thereon—for a low-coverage system where the electronic effects of both the dopant atom and the adspecies are isolated.

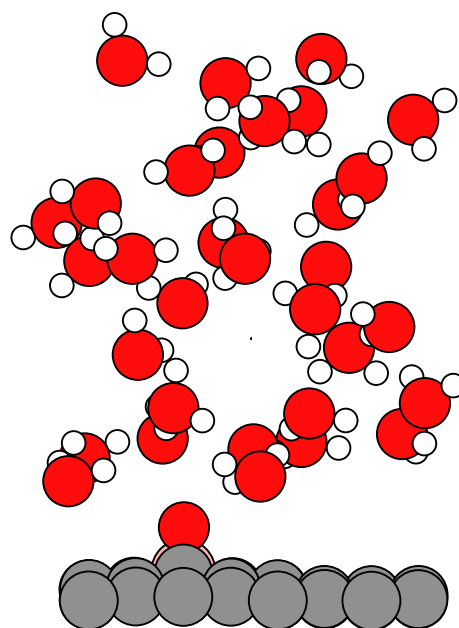


Fig. 1 Rendered illustration of the BG sheet model system with an $*O$ adatom in contact with 32 water molecules (4 layers)

3.2 Simulation Parameters

The obtained data sets, including input files with simulation parameters, are distributed alongside this article and are available from <https://doi.org/10.5281/zenodo.7684918>.

3.3 Choice of DFT Code and Functional

All simulations were performed with the VASP software version 6.2.0 [33–36]. The RPBE density functional [37] with DFT-D3 dispersion correction [15, 16] was used. The RPBE-D3 method has been shown to yield water configurations in agreement with experiments and higher-level methods at comparatively low computational cost [38].

Previous work on NG showed that adsorption energy values for the ORR intermediates can be wrong by up to 0.4 eV compared to the best estimate provided by the HSE06 hybrid functional, which was found to give the lowest error of 5% compared to a diffusion Monte Carlo benchmark calculation [29]. Similar results were obtained for BG [31], see Table S1, where η_{TCM} with the HSE06 functional was *ca.* 1.0 V versus SHE. (Meta-) GGA functionals underestimated this best-estimate value by up to 0.6 V. Figure S3 shows the free energy trends for the ORR on BG obtained with various density functionals. However, our previous work also showed that $\Delta\Delta E_{solv}$ does not share the same strong dependency on the functional [29]. This realization enables the present study since FPMD simulations as long as required for this work are currently not computationally feasible using hybrid functionals.

3.3.1 Static DFT Calculations

Static calculations constitute single-point electronic energy calculations as well as minimization calculations of the total energy with respect to the atomic coordinates. Wave functions were self-consistently optimized until the energy in subsequent iterations changed by less than 10^{-6} eV. The wave function was sampled using Monkhorst–Pack k point grids [39]. A k point density larger than $2 \times 2 \times 1$ was found to give converged results for $\Delta\Delta E_{\text{solv}}$, see Fig. S5. Due to the wide variety of structures calculated in this work, refer to the data set distributed alongside this article to see the chosen k point density for each subset of calculations.

Simulations were carried out using a plane wave basis set with an energy cutoff of 600 eV to represent valence electrons. The projector-augmented wave (PAW) method [40, 41] was used to account for the effect of inner electrons. See Fig. S6 for a convergence study for the PAW energy cutoff. Gaussian-type finite temperature smearing was used to speed up convergence. The smearing width is chosen so that the electronic entropy was smaller than 1 meV in all cases. Real-space evaluation of the projection operators was used to speed up calculations of larger systems, using a precision of 10^{-3} eV atom $^{-1}$. The periodic images are separated by 14 Å of vacuum and a dipole correction is applied perpendicular to the slab.

Atomic coordinates were optimized until the norms of all forces reached below 10^{-2} eV Å $^{-1}$. The L-BFGS limited-memory Broyden optimizer from the VASP Transition State Tools (VTST) software package was used to minimize the forces with respect to the atomic coordinates.

3.3.2 Classical Molecular Dynamics Simulations

Classical molecular dynamics (MD) simulations were carried out in an NVT ensemble at 300 K using the Langevin dynamics [42] implemented in the VASP software. The simulations used similar parameters to those outlined in Sect. 3.3.1 but used a lower PAW energy cutoff of 400 eV and a $3 \times 3 \times 1$ Monkhorst–Pack k point grid to enable total simulation times of up to 100 ps. A Langevin friction parameter of $\gamma = 4.91$ was used throughout all simulations.

Dynamics were calculated initially until the total energy and temperature were converged. This equilibration period is not considered in the evaluation and was optimized on a case-by-case basis. After equilibration had been achieved, the actual sampling was performed over a period of time. In all simulations the geometry of the graphene sheet and the adspecies were constrained to the geometry obtained from a one-shot geometry optimization of the system in contact with $n = 1$ –4 water layers, respectively. Only the water molecules were allowed to move during simulations. The E_{tot}

versus t and T versus t trends for all simulations are shown in the online SI.

Two data sets were generated:

1. First, simulations were performed without any constraints on the water molecules and with a time step of 0.1 fs. Simulations were continued up to a total simulation time of 10 ps after thermalization. This set of MD simulations will be referred to as the *flexible MD* data set going forward.
2. Second, simulations were repeated after placing a Rattle-type bond length constraint [43] on the O–H and H–H bonds to keep the geometry of water molecules rigid throughout simulations, thus enabling a coarse-grained time step of 1.0 fs. Simulations were continued up to a total simulation time of 100 ps after thermalization. This set of MD simulations set will be referred to as the *constrained MD* data set going forward.

To obtain $\Delta\Delta E_{\text{solv}}$, configurations were sampled every 1 ps, yielding 10 samples for the *flexible MD* data set and 100 samples for the *constrained MD* data set. This choice of sampling frequency is informed by the correlation time of water. The correlation time is the time it takes for complete re-orientation of the water arrangement, thus yielding a new, independent sample configuration that is statistically significant. It was found to be *ca.* 1.7 ps for water at room temperature using nuclear magnetic resonance spectroscopy [44]. The chosen sampling rate of 1 ps is smaller than this value as a result of the significant computational effort of performing long dynamics simulations. To minimize the risk of oversampling, Langevin dynamics was chosen to describe coupling to a heat bath. Langevin dynamics introduces a stochastic component to the propagation which can help to diversify configurations more quickly compared to fully deterministic dynamics.

4 Results

4.1 One-Shot Minimization of Atomic Coordinates

The first data set is generated by bringing the BG model system with *O, *OH, and *OOH adspecies into contact with 4–32 molecules of water and minimizing the resulting configurations with respect to the atomic forces. This data set will be referred to as the *one-shot minimization* data set going forward. The water configurations are modeled after those used by Reda et al. to calculate the solvation stabilization energy for the ORR intermediates on NG sheet model systems [23]. Configurations were created so that water molecules are only on one side of the BG sheet

Table 2 Summary of the calculated $\Delta\Delta E_{\text{solv}}$ results from the *one-shot minimization* data set

| # of water molecules | Arrangement | $\Delta\Delta E_{\text{solv}}/\text{eV}$ |
|----------------------|--|--|
| 4 H ₂ O | On side with *O | 0.19 [†] |
| 8 H ₂ O | On side with *O | -0.04 [†] |
| 16 H ₂ O | On side with *O | -0.11 [†] |
| 16 H ₂ O | 8 on both sides | 0.15 [‡] |
| 24 H ₂ O | 16 on side with *O, 8 on the other side | -0.07 [‡] |
| 24 H ₂ O | On side with *O | -0.20 [†] |
| 32 H ₂ O | 24 on side with *O, 8 on the other side | -0.04 [‡] |
| 32 H ₂ O | On side with *O | -0.06 [†] |
| 32 H ₂ O | 16 on side with *O, 16 on the other side | -0.06 |

[†] and [‡] indicate if water molecules are placed only on one side or on both sides of the BG sheet model, respectively

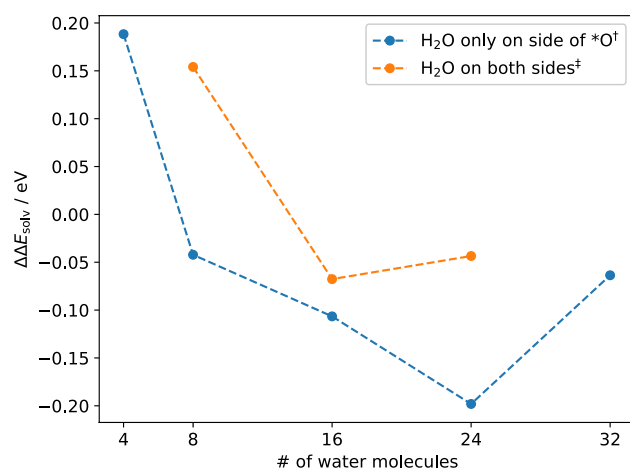


Fig. 2 $\Delta\Delta E_{\text{solv}}$ results for the *O intermediate on BG in contact with 4–32 molecules of water from the *one-shot minimization* data set. The blue line shows $\Delta\Delta E_{\text{solv}}$ for models where water molecules are exclusively placed on the side of the model where the adatom is located. The orange line shows $\Delta\Delta E_{\text{solv}}$ values for select models where water molecules are placed on both sides of the model. For the orange line, the x axis indicates the number of water molecules on the side with the adatom and not the total number of water molecules. The [†] and [‡] indicators connect the values in this figure to the corresponding data values in Table 2

model or on both sides, denoted with the [†] and [‡] symbols, respectively, in Table 2 and Fig. 2.

The $\Delta\Delta E_{\text{solv}}$ results in the *one-shot minimization* data set give rise to several trends. First, when water molecules are placed only on one side of the model, $\Delta\Delta E_{\text{solv}}$ for the *O intermediate does not appear to be converged within the tested series of models as $\Delta\Delta E_{\text{solv}}$ still increases from -0.20 eV to -0.06 from 24 to 32 molecules. Values can be deemed converged if changes are below *ca.* 0.05 eV or 1 kcal mol⁻¹, i.e., chemical accuracy.

Second, the results for simulations where molecules are placed only on the side of the sheet model with the adatom ([†]) are inconsistent with simulations where molecules are placed on both sides of the model ([‡]). For example, deviations of < 0.05 eV are found between simulations where 16 molecules are placed on the side of the adatom and 0, 8, and 16 molecules are placed on the other side. This result would potentially indicate that water molecules on the opposite side of where the adspecies is located have negligible influence and can be omitted. However, the deviation between $\Delta\Delta E_{\text{solv}}$ values where 8 molecules are placed on the side with the adspecies and 0 or 8 molecules are placed on the other side is 0.19 eV. Similarly, the deviation between $\Delta\Delta E_{\text{solv}}$ values where 24 molecules are placed on the side with *O and 0 or 8 molecules are placed on the other side is 0.16 eV.

Results from the *one-shot minimization* data set are therefore inconsistent. From this data, it is unclear if and when $\Delta\Delta E_{\text{solv}}$ will converge as a function of the number of added water molecules and it cannot be assessed with confidence if water molecules do or do not need to be present on the side of the sheet opposite of the adspecies.

One potential reason for the inconsistent behavior lies in the one-shot nature of the data set: water molecule arrangements are flexible and form a complex energy landscape where minimization algorithms can easily become stuck in local minimum configurations. This limitation can be overcome by rigorous sampling of the configurational space using MD integration.

4.2 NVT Simulations

In order to probe if insufficient sampling of the configurational space is responsible for the inconsistent results in the *one-shot minimization* data set, $\Delta\Delta E_{\text{solv}}$ is subsequently determined as an ensemble average by performing MD simulations for a total of 10 ps using a time step of 0.1 fs. No constraint was placed on the O–H and H–H bonds of water molecules. This set of simulations is referred to as the *flexible MD* data set. Due to the significant computational effort of these simulations, only model systems where water molecules are placed on the same side with the adspecies are considered. Simulations are performed for the clean BG sheet model, for the BG sheet with an *O adatom in contact with 8–32 molecules, and for the *OH and *OOH adspecies in contact with 8–24 molecules of water. Figure 3a visualizes the $\Delta\Delta E_{\text{solv}}$ results calculated in this data set.

Focusing on the *O intermediate (blue curve), a similar trend of $\Delta\Delta E_{\text{solv}}$ versus the number of water molecules emerges as before from the *one-shot minimization* data set: the values oscillate and there is an increase of $\Delta\Delta E_{\text{solv}}$ from -0.3 eV to 0.2 eV from 24 to 32 molecules, indicating significant destabilization of this adspecies with increasing number of water molecules.

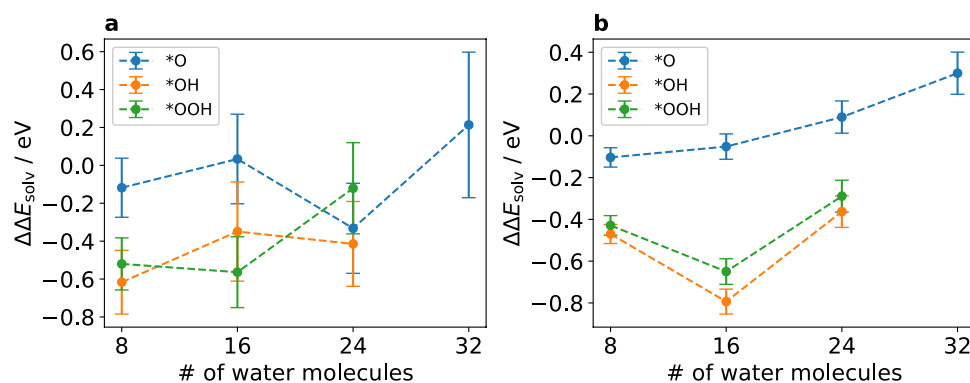


Fig. 3 $\Delta\Delta E_{\text{solv}}$ results for the *O (blue curve), *OH (orange curve), and *OOH (green curve) adspecies on BG in contact with 8–32 molecules of water obtained as ensemble averages from **a** 10 ps of MD using a time step of 0.1 fs where water molecules were flexible and **b**

100 ps of MD using a time step of 1.0 fs where water molecules were constrained. The error bars indicate the two-sided 95% CI calculated according to Eqs. (5)–(8)

It can be summarized that the *flexible MD* data set did not yield more consistent $\Delta\Delta E_{\text{solv}}$ results than the *one-shot minimization* data set. While a similar overall $\Delta\Delta E_{\text{solv}}$ trend is observed for the *O adspecies, differences between subsequent data points are even larger than in the case of the *one-shot minimization* data set.

Another important observation is the size of the error bars, which extend from 0.25 eV up to over 0.5 eV in some cases. Note that in the case of the *O intermediate, the error bar span becomes larger as a function of the number of water molecules. This effect is much less pronounced, if at all observable, for the *OH and *OOH intermediates. However, it is clear from the size of the error bars that the length of simulation time is too short compared to the correlation time of water and thus simulations only yielded 10 independent samples that entered into the evaluation.

In an effort to extend the simulation time, a coarse-graining approach was chosen where the O–H and H–H bond lengths of water molecules were constrained to the average corresponding bond lengths obtained in the *flexible MD* data set. This bond length constraint allows for larger simulation time steps to be taken without the risk of spurious discretization errors from inadequate sampling of the fast O–H vibrations. A subsequent set of dynamics simulations of the same model systems thus used a time step of 1.0 fs and was continued for a total of 100 ps simulation time, yielding 100 independent samples. $\Delta\Delta E_{\text{solv}}$ results from this *constrained MD* data set are visualized in Fig. 3b.

$\Delta\Delta E_{\text{solv}}$ trends from the *constrained MD* data set, while also showing no signs of converging behavior, differ significantly from the *flexible MD* and *one-shot optimization* data sets. The obtained $\Delta\Delta E_{\text{solv}}$ values for the *O adspecies do not oscillate as in the case of the other data sets but continuously increase with increasing number of water molecules. From this data set, the presence of 24 and 32 water molecules is predicted to significantly destabilize

this intermediate. With *ca.* 0.25 eV, the data point for 32 water molecules from this data set is similar to the *flexible MD* data set, however, this data set does not show the reduction of $\Delta\Delta E_{\text{solv}}$ at 24 molecules that was observed for both the *flexible MD* and the *one-shot minimization* data sets.

The *OH and *OOH adspecies show similar $\Delta\Delta E_{\text{solv}}$ trends that parallel each other in this data set; however, values oscillate by up to 0.5 eV when the number of water molecules is increased. Finally, the factor 10 longer simulation time affects the size of the error bars which is now on the scale of *ca.* 0.1 eV. Similar to results from the *flexible MD* data set, the error bars for $\Delta\Delta E_{\text{solv}}$ of the *O adspecies are found to increase with increasing number of water molecules in the simulation while no such trend is observed for the *OH and *OOH intermediates.

Finally, the local structure of the water molecules around the adspecies is analyzed using z distribution functions, $g(z)$, shown in Fig. S7. The $g(z)$ distributions are obtained by calculating the distances between the O atoms of water molecules and an $x - y$ plane located at the average z coordinate of the atoms in the BG sheet model. The $g(z)$ show distinct bands for the first and second solvation layer. The bands for the third and fourth layers are significantly more broadened, indicating that the surface-adjacent water double layer is more strongly coordinated compared to subsequent layers. Notably, shoulders at the first band are visible in the $g(z)$ generated from the *flexible MD* data set which are not visible in those generated from *constrained MD* data set. However, this result is presented with the caveat that the data is more noisy compared to the smoother *constrained MD* $g(z)$ results due to the 10× smaller sampling statistics. This result potentially indicates that the bond length constraint affects the coordination fine structure around the adspecies and thus may help to explain the differences between the *flexible MD* and *constrained MD* data sets. However, more

detailed investigation is required to validate the importance of this observed difference.

It can be summarized that coarse-grained MD simulations yielded a data set that is significantly different from the more similar-to-each-other *flexible MD* and *one-shot minimization* data sets but did not yield more consistent $\Delta\Delta E_{\text{solv}}$ results overall. Finally, the bond length constraint is found to change the $\Delta\Delta E_{\text{solv}}$ results compared to the *flexible MD* data set; however, since there are currently no converged reference values for $\Delta\Delta E_{\text{solv}}$, it is impossible to assess if the changes introduced by the Rattle-type constraint are detrimental to the results or not.

4.3 Re-sampling and Energy Minimization

The *flexible MD* and *constrained MD* data sets did not yield converged $\Delta\Delta E_{\text{solv}}$ results. There are, however, two technical limitations which may reduce the significance of these data sets:

1. For these data sets, $\Delta\Delta E_{\text{solv}}$ is calculated by using the average total energy from an *NVT* ensemble ($T = 300$ K) for the energy terms labeled “with solvent” in equation (4). The energy terms labeled “without solvent” are obtained from static energy minimization calculations of the systems without solvent which are technically at 0 K temperature. While the BG sheet model and adspecies were kept frozen in the atomic configuration from a 0 K energy minimization during the MD and only water molecules were allowed to move, it cannot be fully excluded that results are biased due to a mismatch between the averaged finite-temperature MD values on one side and the locally optimized, 0 K values on the other side of the equation.
2. As outlined in Sect. 3, the MD simulations—as well as the corresponding reference simulations of the systems “without solvent” needed for Eq. (4)—used a reduced PAW energy cutoff value of 400 eV to enable longer simulation times. This value is technically not converged for adsorption energy calculations, see Fig. S5.

In order to address both of these limitations, a fourth data set is produced. To this end, 20 structures are randomly sampled from each *flexible MD* trajectory and subsequently energy-minimized using the settings presented in Sect. 3, i.e., with a larger PAW energy cutoff of 600 eV. This way, the diversity of the MD-generated configurations is maintained but all values entering Eq. (4) are obtained from energy-minimized atomic configurations using safer accuracy settings. This data set will be referred to as the *resampled* data set going forward. Figure 4 visualizes the $\Delta\Delta E_{\text{solv}}$ results from this data set.

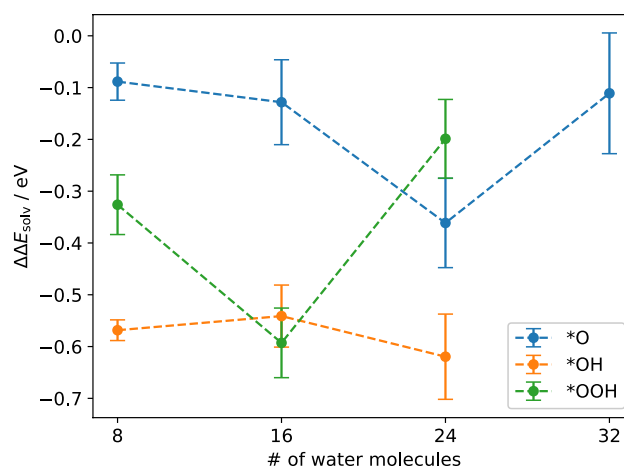


Fig. 4 $\Delta\Delta E_{\text{solv}}$ results for the *O, *OH, and *OOH adspecies on BG in contact with 8–32 molecules of water obtained as average values over 20 images per data point which were randomly resampled from the flexible MD data set and subsequently energy-minimized with respect to the atomic coordinates. The error bars indicate the two-sided 95% CI calculated according to Eqs. (5)–(8)

The *resampled* data set shares similarities with the *flexible MD* and *one-shot optimization* data sets, for example the characteristic dip of $\Delta\Delta E_{\text{solv}}$ for the *O adatom at 24 water molecules. This result further indicates that the bond length constraint used to obtain the *constrained MD* data set is likely altering the trends in a significant way. The previously discussed trend regarding error bar spans increasing with increasing number of molecules is distinctly present both for the *O and the *OH adspecies. Ultimately, this data set does not provide fundamentally different insights into the $\Delta\Delta E_{\text{solv}}$ trends compared to the preceding analyses.

5 Discussion

5.1 Comparison of the Results from Different Data Sets

Figure 5 shows a side-by-side comparison of $\Delta\Delta E_{\text{solv}}$ as a function of the number of water molecules for the *O, *OH, and *OOH adspecies from the four obtained data sets.

The *resampled* data set is the most significant data set among those obtained in this work as it combines the broad configurational diversification of the MD simulations with the methodological consistency of calculating $\Delta\Delta E_{\text{solv}}$ using strict accuracy parameters and exclusively on the basis of energy-minimized structures. By comparing the data sets with each other and with the *resampled* data set in particular, several important aspects can be highlighted.

First, convergence of $\Delta\Delta E_{\text{solv}}$, i.e. changes of < 0.05 eV between subsequent data points, is not observed in any case.

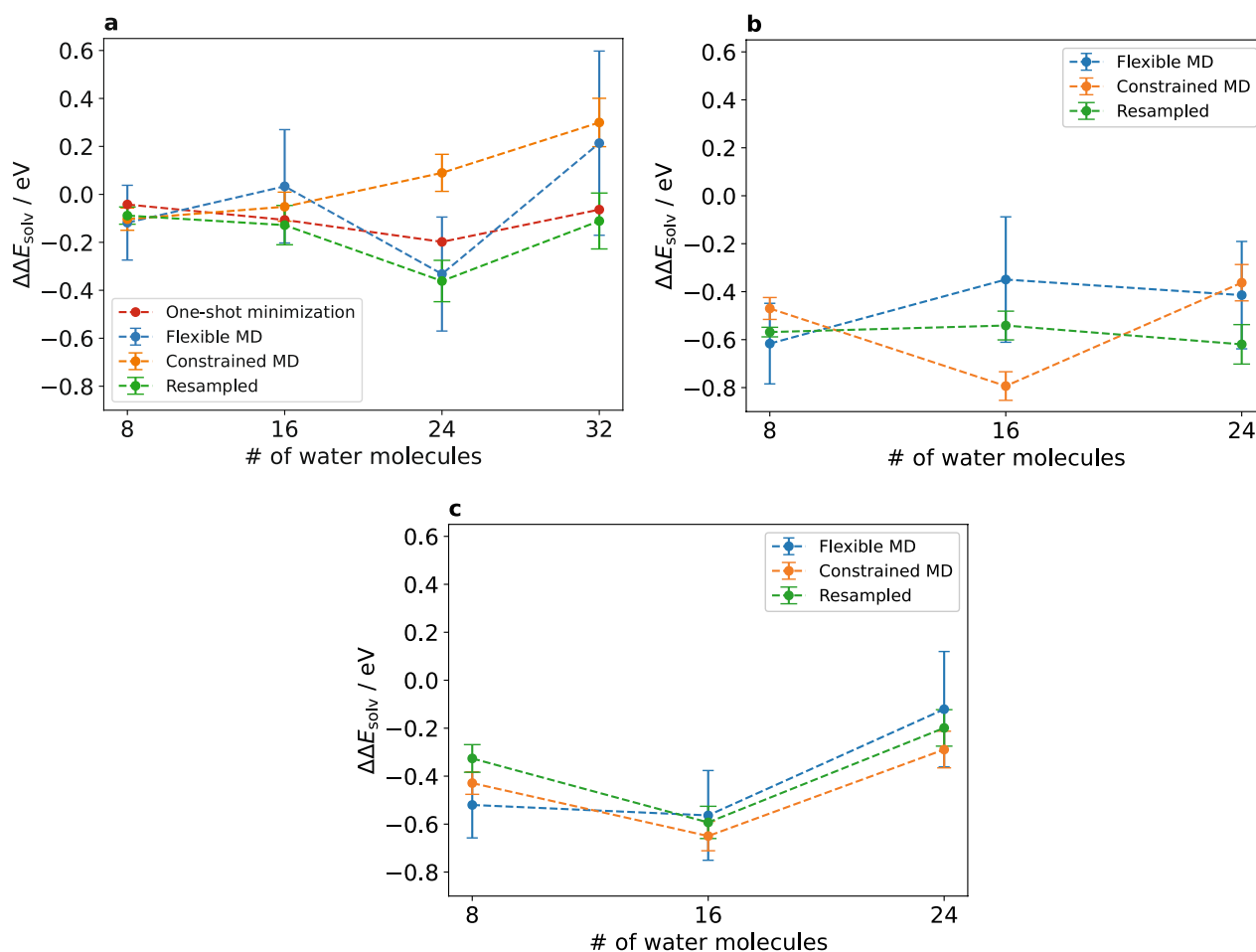


Fig. 5 Comparison of $\Delta\Delta E_{\text{solv}}$ results for the **a** *O, **b** *OH, and **c** *OOH adspecies from the *one-shot minimization* data set, the *flexible MD* and *constrained MD* data sets, and the *resampled* data set

It is impossible at this point to give a confident estimate of $\Delta\Delta E_{\text{solv}}$ for the tested adspecies on the BG sheet model. This result indicates that more than 32 molecules (4 layers) of water are necessary to obtain converged results.

Converging the $\Delta\Delta E_{\text{solv}}$ value to changes within chemical accuracy is of crucial importance. For example, consider the potential-dependent free energy trends for the ORR on the BG model presented in Fig. S3. These trends were obtained according to the free energy approach using the computational hydrogen electrode [13]. Using the most reliable functional for adsorption energy calculations on this material class according to benchmarks [29, 30, 45], the HSE06 hybrid functional, the potential-determining step is the formation of the *OOH intermediate by a significant margin. The extrapolated thermochemical overpotential, η_{TCM} , for the ORR on the present BG model is *ca.* 1.0 V versus SHE. Stabilization of the *OOH intermediate by roughly -0.4 eV (8 water molecules), -0.6 eV (16 water molecules), or -0.2 eV (24 water molecules) will therefore proportionally reduce

η_{TCM} to 0.6 V, 0.4 V, and 0.8 V versus SHE, respectively. Therefore, depending on the number of included water molecules, one can predict a mostly inactive ($\eta_{\text{TCM}} = 0.8$ V, 24 molecules) or moderately active ($\eta_{\text{TCM}} = 0.4$ V, 16 molecules) ORR electrocatalyst. The overpotential of a typical reference Pt/C electrocatalyst is 0.3–0.4 V [18]. Therefore, $\Delta\Delta E_{\text{solv}}$ must be converged within the limits of chemical accuracy before any trustworthy prediction can be made.

Second, there appears to be no obvious systematicity to whether trends from the different data sets agree with each other or not. For example, values from different data sets for the *OOH intermediate are in reasonable agreement and show similar overall trends. In the case of the *O adatom, there is some correlation between trends from correlated data sets (in particular the *flexible MD* data set and the *resampled* data set which was generated from the former) and only the *constrained MD* data set behaves significantly different. In the case of *OH, however, there appears to be no shared trends between results from either of the data sets.

Further research is needed to analyze why there is reasonable agreement in some cases and no agreement in other cases.

Third, the error bars in all cases are significantly larger than chemical accuracy (± 0.05 eV). Aside from the fluctuation amplitude of the total energy values, the size of the error bar is governed by the number of independent samples. Because of the long experimentally measured correlation time of water, significantly longer statistics may be required to reduce the uncertainty to within chemical accuracy. See also Sect. 5.3.1 for a detailed analysis of the influence of sampling frequency.

Fourth, from the results presented in Table 2, it cannot be completely ruled out that water molecules may have to be added to both sides of the BG sheet model to obtain correct results. This result stands in contrast to results by Reda et al. for NG where results for placing water molecules on one side or both sides of the model were close to identical [23]. This result therefore shows that $\Delta\Delta E_{\text{solv}}$ values obtained for one material cannot be transferred to others, even if they are as closely related as NG and BG.

Fifth, analysis of the z distributions, $g(z)$, of oxygen atoms from the water molecules based on the MD data sets provided some first evidence that the bond length constraint used to obtain the *constrained MD* data set may have affected the coordination fine structure around the adspecies. However, due to the poor statistics resulting from the small required time step used to generate the *flexible MD* data set, it would be necessary to extend these simulations by a factor 5–10 to obtain enough independent samples to make sure that this observation is significant.

As an intermediary conclusion, the most likely explanation for the non-convergence of the $\Delta\Delta E_{\text{solv}}$ results in general, as well as for the non-systematic differences between data sets more specifically, is that significantly more water molecules need to be included in simulations. It is unclear at this point how many water molecules would be required to achieve convergence. Sakong et al. found that 6 layers of water are needed to obtain bulk water behavior and converged work function estimates in the case of FPMD simulations of a Pt(111) surface in contact with water [46]. However, Pt(111) is a strongly-coordinating surface compared to the hydrophobic BG sheet model in the present study. Furthermore, the group tested for convergence of the work function and not for $\Delta\Delta E_{\text{solv}}$ of reaction intermediates. Hence, it is unlikely that the number of 6 necessary water layers will also be the correct number of layers to include for the present system.

For these reasons, it is currently not possible to foresee the ultimately required number of water molecules required to obtain converged $\Delta\Delta E_{\text{solv}}$ results for this system. Attempting to find this number systematically by dynamics simulations with DFT atomic forces quickly becomes computationally

unfeasible; simulations for the models in contact with 32 water molecules in this work already required several weeks of computational time. Even if these considerable time and energy resources would be spent to identify the required number of water molecules for the present problem, such a study would have to be repeated for every new material under investigation. Even though the influence of solvation has been shown to significantly affect free energy trends, the authors are therefore convinced that such simulations cannot yet be performed routinely.

We have thus come to the decision to publish the present results as-is and to not continue simulations with model systems that include more and more water molecules at ever increasing computational cost. Instead, we are currently focusing research efforts into development of a 2D-periodic polarizable-embedding QMMM method that will allow for simulations with thousands of water molecules while retaining electronic structure level accuracy for the surface model and the closest few layers of water molecules. This method will use the Single Center Multipole Expansion (SCME) ansatz to describe polarization of water molecules which is crucial to accurately describe interface processes such as charge transfer [47, 48]. Because the boundary plane between the QM and MM regions has exclusively water molecules on both sides, and because it is not necessary to describe diffusion to or from the surface to obtain $\Delta\Delta E_{\text{solv}}$ results, an efficient restrictive boundary method can be used. The SAFIRES method recently developed in our groups was built to support 2D periodic boundary conditions [49].

A publication on the technical implementation of the 2D periodic polarizable-embedding QMMM ansatz for the open-source GPAW and ASE programs is currently in preparation in our groups. The goal is to use this method to revisit the BG model system in the present work.

5.2 Comparison to Literature Results

To the best of our knowledge, there is only one other study in literature where $\Delta\Delta E_{\text{solv}}$ values from explicit solvation were calculated for the ORR intermediates on BG. Fazio et al. used a molecular BG flake model in contact with a cluster of 6 water molecules to obtain $\Delta\Delta E_{\text{solv}}$ [17]. The group used the B3LYP hybrid functional in combination with DFT-D3 dispersion correction. From this model, they obtained $\Delta\Delta E_{\text{solv}}$ values of -0.06 eV, -0.37 eV, and -0.46 eV for the *O, *OH, and *OOH intermediates. The values for *O and *OOH are in reasonable agreement with the results for 8 water molecules in the present study, which is the closest point of reference. The value for *OH is 0.15 to 0.20 eV more positive than in the present work. Because the $\Delta\Delta E_{\text{solv}}$ values in the present work are not converged even when 32 water molecules are included, an

in-depth discussion about potential reasons for the (dis-) agreement of the present results and the results by Fazio et al. is not appropriate.

However, results for $\Delta\Delta E_{\text{solv}}$ for the ORR intermediates on NG obtained with periodic model systems and a larger number of water molecules are available. We will therefore attempt to compare as best as possible with the results of the closely related NG system.

The *one-shot minimization* data set was calculated in a similar way as free energy results presented by Reda et al. for the ORR intermediates on NG [23]. In the case of Reda et al., the solvation energy results were found to be converged when one layer of water molecules was included, and including water molecules on both sides of the surface was not found to strongly impact results. It is currently not clear why the present results for BG show such different trends.

One potential cause for this disparity could be the slightly different binding geometry of the ORR intermediates on NG and BG. The *O intermediate is bridge-bound for BG and bound on a C-top position for NG and the *OH and *OOH intermediates are bound on the B-top position on BG and on a C-top position on NG. However, it is unclear if these arguably subtle differences are responsible for the difference in convergence trends for $\Delta\Delta E_{\text{solv}}$. Another potential cause for the disparity may be the simulation approach chosen by Reda et al. who used a global minimization algorithm to find the optimal H₂O arrangements. Finally, Reda et al. used the BEEF-vdW functional while RPBE-D3 was used in the present work. Detailed benchmarks would be required to establish if the different density functional or dispersion method cause the diverging behavior.

Another comparison can be made with FPMD results for the ORR on NG presented by Yu et al. [21]. The group estimated $\Delta\Delta E_{\text{solv}}$ by introducing 41 water molecules to a NG model, performing classical dynamics simulations with DFT forces, and finally minimizing the lowest-energy solvated structures obtained from the MD simulation with respect to the atomic coordinates. The group obtained $\Delta\Delta E_{\text{solv}}$ values of -0.53 , -0.38 , and -0.49 eV for the *O, *OH, and *OOH intermediates, respectively.

While this approach fails to capture the vast structural diversity of the configurational space and is therefore less representative of the system under experimental conditions, it has value from a computational perspective because $\Delta\Delta E_{\text{solv}}$ according to Eq. 4 is calculated exclusively from 4 values total, all of which represent the best possible guess for the global minimum energy configuration of each system.

Hence, we also apply this approach to the present data set to check if $\Delta\Delta E_{\text{solv}}$ trends become less erratic by this way of analysis. The *flexible MD* data set was re-analyzed to find the structure with the lowest total energy for each combination of adspecies and number of water molecules. The obtained images are then energy-minimized using the safe accuracy

settings outlined in Sect. 3. Figure S8 shows the results of this approach.

Figure S8 shows that the $\Delta\Delta E_{\text{solv}}$ results for *OH and *OOH are comparable to the *resampled* data set in terms of relative trends but less so in terms of absolute values. However, the *O intermediate shows significantly more negative $\Delta\Delta E_{\text{solv}}$ results.

It can therefore be concluded that this approach not only did not resolve the erratic results but can further distort the results because the close-to-ideal local configurations optimized in this case likely do not represent the average configurations of water molecules around the adspecies in real, finite-temperature systems.

5.3 Analysis of Potential Error Sources

To conclude the discussion of the data sets presented in this work, the following sections will rule out various potential error sources that readers familiar with dynamics simulations and the pitfalls of solvation energy calculations may be concerned about. The obtained simulation data sets and data evaluation workflow have also been made available online, see section Supplementary Information.

5.3.1 Influence of the Sampling Frequency on the Results

Configurations were sampled from the dynamics simulations at an interval of 1 ps. It is important to ask how the $\Delta\Delta E_{\text{solv}}$ results are affected by changes of the sampling frequency. Figure S9 compares $\Delta\Delta E_{\text{solv}}$ results from the *flexible MD* and *constrained MD* data sets analyzed every 2 ps, 1 ps, 100 fs, and 10 fs.

The $\Delta\Delta E_{\text{solv}}$ results appear to be robust against the choice of sampling frequency. The only significant differences are observed between the *flexible MD* data set sampled every 2 ps (5 total samples) and sampled every 1 ps (10 total samples) and faster. This difference can be attributed to the poor statistics in the case of the 2 ps sampling frequency.

The size of the error bars is affected significantly by the sampling frequency because the square root of the number of samples, \sqrt{n} , enters the divisor of Eq. (7). This test therefore highlights the importance of choosing a reasonable sampling frequency based on the physical properties of the system to obtain a meaningful error bar. It is easy to get lured into a false sense of security by oversampling the results to obtain small error bars.

5.3.2 Spurious Dipole and Quadrupole Corrections

Total energy calculations were performed using dipole and quadrupole correction perpendicular to the surface to avoid interactions between periodic repetitions of the simulation box. It is known that first-row semiconductors with defects,

of which BG is an example, can lead to large dipole and quadrupole moments, thus making the correction necessary. However, our simulations showed that the correction can sometimes give erroneously large corrections of several eV for unknown reasons. After re-optimizing the wave function in a single-point calculation, the correction is then found to be of a reasonable magnitude again, usually on the order of some meV.

Because it is impossible to perform this manual correction for all calculations in this work, the consistency of the results is representatively examined by analyzing the average dipole and quadrupole correction energy (and uncertainty thereof) of the *resampled* data set. Figure S10 shows the results of this analysis. The average correction energy is ≤ 0.02 eV in all cases, which is within the limits of chemical accuracy. Error bars are found to be as large as 0.01 eV in some cases and close to 0.02 eV in one extreme case (BG-OOH in contact with 24 water molecules), indicating that the dipole and quadrupole energy correction is indeed volatile (in relation to the absolute values) and dependent on the exact geometry of the system. However, due to the small overall magnitude of the correction, it can be concluded that this correction should not significantly influence the calculation results.

5.3.3 Spurious Dispersion Correction

DFT-D3 dispersion correction values are significantly larger in magnitude than the dipole and quadrupole correction energy discussed in Sect. 5.3.2. Figure S11a uses the *resampled* data set to show the dispersion energy difference $\Delta E_{\text{disp}} = E_{\text{disp}}^{\text{BG-adspecies}} - E_{\text{disp}}^{\text{BG-clean}}$ between the BG systems with the adspecies *O, *OH, and *OOH and the clean system, all of which are in contact with water. This analysis therefore highlights the contribution of the dispersion energy to the adsorption energy for the solvated model systems. Figure S11b reproduces the $\Delta\Delta E_{\text{solv}}$ results as a function of the number of water molecules shown in Fig. 4 but with the dispersion energy removed from the total energy.

This analysis shows that the dispersion contributions increase with the size of the solute. ΔE_{disp} is close to zero for the *O adatom but *ca.* -0.5 eV for *OOH in contact with 16 water molecules. The values for *OH and *OOH fluctuate significantly between subsequent data points, raising the question if the dispersion correction may be partially responsible for the erratic behavior of the $\Delta\Delta E_{\text{solv}}$ trends. However, analyzing the $\Delta\Delta E_{\text{solv}}$ trends in Fig. S11b shows that the results do not become more consistent when the dispersion energy contribution is removed. Hence, it can be concluded that any volatility of the dispersion correction results is also not the cause for but most likely the result of the erratic nature of the entire data set.

One caveat in this analysis and discussion, however, is that this *a posteriori* removal of the final dispersion correction energy does not remove the entire influence of dispersion correction on the data set. Both the MD simulations and the local minimization of the structures in the *resampled* data used dispersion correction throughout, hence the final structures (re-)analyzed here are generated on the RPBE-D3 potential surface. Despite this caveat, it is still unlikely that dispersion is the driving factor behind the erratic results since in particular the RPBE-D3 functional combination has been shown in the past to produce water structure that is in good agreement with experiments [38].

5.3.4 Influence of Simulation Cell Size

Simulation cells varied in size between simulations with different number of included water molecules. Because a PAW-based DFT approach was used and PAWs always fill the entire simulation cell, the *c* cell parameter was minimized on a case-by-case basis to minimize the computational effort. Increasing or decreasing the box size also changes the total energy in a small way, hence it is important that all energy values used to calculate $\Delta\Delta E_{\text{solv}}$ in Eq. (4) use the same cell dimensions. Consistency in this regard was ensured by generating the reference systems without solvent by removing water molecules from the original system; the reference systems are given alongside the solvated parent models in the data set available from <https://doi.org/10.5281/zenodo.7684918>.

Furthermore, Table S2 summarizes the total energy results for various reference systems without solvent from the MD data sets. The differences between system are, despite differences in the *c* cell parameter, < 0.01 eV. Hence, the total energy contributions from inconsistent cell dimensions, even if they had been left untreated, are unlikely to distort results enough to account for the erratic results in this work.

5.3.5 Influence of Minimizing the Reference Systems

This concern is related to the discussion about inconsistent cell size in Sect. 5.3.4. As pointed out there, the reference systems were obtained from the solvated parent systems by removal of the water molecules and subsequent energy-minimization of the resulting atomic configurations. This approach was chosen to account for the possibility that the most stable atomic arrangement of the BG-adspecies system may change once water molecules are removed.

However, this approach creates a potential inconsistency: by optimizing the atomic configuration of the reference systems, the $\Delta\Delta E_{\text{solv}}$ values obtained from Eq. (4) do not only contain the interaction of the BG-adspecies system with the water molecules but also the reorganization energy of the

systems when going from a system in vacuum to a solvated system.

To investigate if energy minimization of the atomic configuration of the reference systems creates a bias, Fig. S12 compares $\Delta\Delta E_{\text{solv}}$ results from the *one-shot optimization* data set where the reference systems without solvent were either minimized or where the reference energy contributions $E_{\text{tot}}^{\text{BG+adatom without solvent}}$ and $E_{\text{tot}}^{\text{BG without solvent}}$ were obtained from single-point total energy calculations.

Results from this test show that the overall trends are identical. However, $\Delta\Delta E_{\text{solv}}$ for the adspecies in contact with 16, 24, and 32 water molecules are *ca.* 0.2 eV more negative when obtained from single-point energy calculations based on the formerly-solvated atomic configurations. This result is unsurprising because the reference systems without water molecules can be assumed to be in a slightly unfavorable configuration when not allowed to relax under the new environmental conditions.

Overall, the differences appear to be systematic across the board and do not change the trends. Therefore, this factor is also not responsible for the erratic, non-converging behavior of $\Delta\Delta E_{\text{solv}}$ with increasing number of water molecules.

5.3.6 Influence of Constraining the Geometry of the BG Sheet

100 ps of classical dynamics without bond length constraints on the water molecules and *no* geometry constraint on the BG sheet and adspecies were accidentally performed for the BG-OOH system in contact with 1 layer of water. This mistake, however, can be used to probe the influence of the geometry constraint on the BG-OOH system.

Figure S13 compares the total energy and temperature trends over the course of the simulation time for the simulations with and without geometry constraint on the BG-OOH system. Most notably, the total energy fluctuations are significantly increased in the case of the model without constraint. The increased amplitude of fluctuations translate into a larger error bar. Hence, without the geometry constraint on the BG-OOH backbone, more sampling statistics is required to reduce the uncertainty to an appropriate level. In the interest of computational feasibility, the geometry constraint is therefore critical.

Finally, Fig. S14 compares the $g(z)$ of the systems where the BG sheet was constrained against that of the non-constrained system. No significant differences were observed. This result indicates that constraining the BG sheet does not significantly affect the interactions between the surface and the first water layer from a structural point of view.

5.3.7 Embedded Solvation Approach

The embedded solvation approach, where a small cluster of explicit solvent molecules is used in combination with an implicit continuum description of the solvent bulk, has recently been employed to good effect [50, 51]. In the beginning of this study, the *one-shot minimization* data set was, in fact, computed using the embedded approach and similarly erratic results were obtained. The implicit solvent model was then discarded for the remainder of this study to reduce the number of potential error sources.

The key issue with the embedded approach, as highlighted recently by Basdogan et al. [52] is that the number of included explicit water molecules cannot be chosen arbitrarily but must be optimized, for which the group suggests a machine-learned model [53]. In fact, including too many explicit molecules can lead to diverging errors [53]. In the case of graphene models, one would also necessarily have to include water molecules on both sides of the model to avoid inconsistencies. Otherwise, the non-solvated side of the system will give rise to a graphene implicit-bulk-water interaction energy while the solvated side gives the desired water implicit-bulk-water interaction. This increases computational effort again.

While it is certainly conceivable that the approach could be optimized for this system—although there are only few reports [52] so far of this approach being used for periodic surface models—it bears the inherent frustration that the approach will need to be optimized over again for every new model system, which limits the comparability of different models. With a fully explicit approach, on the other hand, one could identify the largest number of water molecules needed for any system in a portfolio and perform all simulations consistently using the same settings.

6 Conclusion

Density functional theory-driven minimization calculations and classical molecular dynamics simulations were used to obtain the solvation stabilization energy, $\Delta\Delta E_{\text{solv}}$, for the oxygen reduction reaction intermediates *O, *OH, and *OOH adsorbed on a Boron-doped graphene sheet in contact with 8, 16, 24, and 32 molecules of water. The goal of this study was to apply the obtained $\Delta\Delta E_{\text{solv}}$ values to accurate hybrid DFT adsorption energy results for the ORR intermediates to refine potential-dependent free-energy predictions. Although 4 different data set were obtained that sampled $\Delta\Delta E_{\text{solv}}$ from the model systems in different ways using static and dynamic calculations, no converged $\Delta\Delta E_{\text{solv}}$ result were obtained.

A detailed discussion of the simulation parameters and potential error sources is provided to rule out that technical

errors lead to these erratic results. We conclude that 32 water molecules, which is the equivalent of 4 layers of water in this model system, are not sufficient to describe solvation of the adspecies within chemical accuracy. Achieving chemical accuracy, i.e. convergence of $\Delta\Delta E_{\text{solv}}$ to changes of ≤ 0.05 eV when adding more and more water molecules, is essential since any reduction of the free energy of the potential-determining intermediate will lead to a proportional reduction of the predicted thermochemical overpotential.

These results emphasize that new simulation methods are required to be able to calculate large enough systems to obtain converged $\Delta\Delta E_{\text{solv}}$ results since molecular dynamics simulations with DFT forces quickly become computationally unfeasible when adding more and more water molecules. Our groups are therefore focused on implementing a 2D-periodic hybrid method (often referred to as QM/MM) for the open-source ASE and GPAW software packages which will enable calculations with thousands of water molecules.

Another promising approach to tackle this problem is the recently developed on-the-fly machine learning force field training method [54]. This approach could be used to train a machine learning force field on a small system and then upscale the system to contain many water molecules while retaining close-to-DFT accuracy.

Finally, we believe in the importance of presenting these negative results to the catalysis community as a word of caution. It is easy to underestimate the number of explicit water molecules required to obtain sufficiently accurate solvation energy results.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11244-023-01829-0>.

Acknowledgements This work was supported in part by the Icelandic Research Fund (Grant No. 174582-053).

Author Contributions All authors contributed to the study conception and design. Simulations, data collection and analysis were performed by BK. The first draft of the manuscript was written by BK and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported in part by the Icelandic Research Fund (grant no. 174582-053). Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Wang H, Maiyalagan T, Wang X (2012) Review on recent progress in nitrogen-doped graphene: synthesis, characterization, and its potential applications. *ACS Catal* 2(5):781–794. <https://doi.org/10.1021/cs200652y>
2. Zhang J, Xia Z, Dai L (2015) Carbon-based electrocatalysts for advanced energy conversion and storage. *Sci Adv*. <https://doi.org/10.1126/sciadv.1500564>
3. Agnoli S, Favaro M (2016) Doping graphene with boron: a review of synthesis methods, physicochemical characterization, and emerging applications. *J Mater Chem A* 4(14):5002–5025. <https://doi.org/10.1039/C5TA10599D>
4. Qu L, Liu Y, Baek J-B, Dai L (2010) Nitrogen-doped graphene as efficient metal-free electrocatalyst for oxygen reduction in fuel cells. *ACS Nano* 4(3):1321–1326. <https://doi.org/10.1021/nn901850u>
5. Sheng Z-H, Gao H-L, Bao W-J, Wang F-B, Xia X-H (2011) Synthesis of boron doped graphene for oxygen reduction reaction in fuel cells. *J Mater Chem* 22(2):390–395. <https://doi.org/10.1039/C1JM14694G>
6. Hummers WS, Offeman RE (1958) Preparation of graphitic oxide. *J Am Chem Soc* 80(6):1339. <https://doi.org/10.1021/ja01539a017>
7. Marcano DC et al (2010) Improved synthesis of graphene oxide. *ACS Nano* 4(8):4806–4814. <https://doi.org/10.1021/nn1006368>
8. Wang L, Ambrosi A, Pumera M (2013) metal-free catalytic oxygen reduction reaction on heteroatom-doped graphene is caused by trace metal impurities. *Angew Chem Int Ed* 52(51):13818–13821. <https://doi.org/10.1002/anie.201309171>
9. Masa J, Xia W, Muhler M, Schuhmann W (2015) On the role of metals in nitrogen-doped carbon electrocatalysts for oxygen reduction. *Angew Chem Int Ed* 54(35):10102–10120. <https://doi.org/10.1002/anie.201500569>
10. Ambrosi A et al (2012) Chemically reduced graphene contains inherent metallic impurities present in parent natural and synthetic graphite. *Proc Natl Acad Sci USA* 109(32):12899–12904. <https://doi.org/10.1073/pnas.1205388109>
11. Xu X et al (2014) Facile synthesis of boron and nitrogen-doped graphene as efficient electrocatalyst for the oxygen reduction reaction in alkaline media. *Int J Hydrog Energy* 39(28):16043–16052. <https://doi.org/10.1016/j.ijhydene.2013.12.079>
12. Jiao Y, Zheng Y, Jaroniec M, Qiao SZ (2014) Origin of the electrocatalytic oxygen reduction activity of graphene-based catalysts: a roadmap to achieve the best performance. *J Am Chem Soc* 136(11):4394–4403. <https://doi.org/10.1021/ja500432h>
13. Nørskov JK et al (2004) Origin of the overpotential for oxygen reduction at a fuel-cell cathode. *J Phys Chem B* 108(46):17886–17892. <https://doi.org/10.1021/jp047349j>
14. Wang L et al (2016) Potential application of novel boron-doped graphene nanoribbon as oxygen reduction reaction catalyst. *J Phys Chem C* 120(31):17427–17434. <https://doi.org/10.1021/acs.jpcc.6b04639>
15. Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate ab initio parametrization of density functional

- dispersion correction (DFT-d) for the 94 elements h-pu. *J Chem Phys* 132(15):154104. <https://doi.org/10.1063/1.3382344>
16. Grimme S, Ehrlich S, Goerigk L (2011) Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* 32(7):1456–1465. <https://doi.org/10.1002/jcc.21759>
 17. Fazio G, Ferrighi L, Valentin CD (2014) Boron-doped graphene as active electrocatalyst for oxygen reduction reaction at a fuel-cell cathode. *J Catal* 318:203–210. <https://doi.org/10.1016/J.JCAT.2014.07.024>
 18. Vielstich W, Lamm A, Gasteiger H (eds) (2003) *Handbook of fuel cells: fundamentals, technology and applications*. Wiley, Chichester
 19. Ferrighi L, Datteo M, Di Valentin C (2014) Boosting graphene reactivity with oxygen by boron doping: density functional theory modeling of the reaction path. *J Phys Chem C* 118(1):223–230. <https://doi.org/10.1021/jp410966r>
 20. Okamoto Y (2009) First-principles molecular dynamics simulation of O₂ reduction on nitrogen-doped carbon. *Appl Surf Sci* 256(1):335–341. <https://doi.org/10.1016/j.apsusc.2009.08.027>
 21. Yu L, Pan X, Cao X, Hu P, Bao X (2011) Oxygen reduction reaction mechanism on nitrogen-doped graphene: a density functional theory study. *J Catal* 282:183–190. <https://doi.org/10.1016/J.JCAT.2011.06.015>
 22. Chai G-L, Hou Z, Shu D-J, Ikeda T, Terakura K (2014) Active sites and mechanisms for oxygen reduction reaction on nitrogen-doped carbon alloy catalysts: stone-wales defect and curvature effect. *J Am Chem Soc* 136(39):13629–13640. <https://doi.org/10.1021/ja502646c>
 23. Reda M, Hansen HA, Vegge T (2018) DFT study of stabilization effects on *n*-doped graphene for ORR catalysis. *Catal Today* 312:118–125. <https://doi.org/10.1016/J.CATTOD.2018.02.015>
 24. Cramer CJ, Truhlar DG (1999) Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chem Rev* 99(8):2161–2200. <https://doi.org/10.1021/cr960149m>
 25. Skyner RE, McDonagh JL, Groom CR, Mourik TV, Mitchell JBO (2015) A review of methods for the calculation of solution free energies and the modelling of systems in solution. *Phys Chem Chem Phys* 17(9):6174–6191. <https://doi.org/10.1039/C5CP00288E>
 26. Zhang J, Zhang H, Wu T, Wang Q, van der Spoel D (2017) Comparison of implicit and explicit solvent models for the calculation of solvation free energy in organic solvents. *J Chem Theory Comput* 13(3):1034–1043. <https://doi.org/10.1021/acs.jctc.7b00169>
 27. Gray CM, Saravanan K, Wang G, Keith JA (2017) Quantifying solvation energies at solid/liquid interfaces using continuum solvation methods. *Mol Simul* 43(5–6):420–427. <https://doi.org/10.1080/08927022.2016.1273525>
 28. Heenen HH, Gauthier JA, Kristoffersen HH, Ludwig T, Chan K (2020) Solvation at metal/water interfaces: an ab initio molecular dynamics benchmark of common computational approaches. *J Chem Phys* 152(14):144703. <https://doi.org/10.1063/1.5144912>
 29. Kirchhoff B et al (2021) Assessment of the accuracy of density functionals for calculating oxygen reduction reaction on nitrogen-doped graphene. *J Chem Theory Comput* 17:6405–6415. <https://doi.org/10.1021/acs.jctc.1c00377>
 30. Hsing CR, Wei CM, Chou MY (2012) Quantum Monte Carlo investigations of adsorption energetics on graphene. *J Phys Condens Matter* 24(39):395002. <https://doi.org/10.1088/0953-8984/24/39/395002>
 31. Kirchhoff B (2021) *Computational studies of oxygen reduction catalysts*, PhD dissertation, Faculty of Physical Sciences, University of Iceland, p 178. <https://hdl.handle.net/20.500.11815/2594>
 32. Grossfield A et al (2019) Best practices for quantification of uncertainty and sampling quality in molecular simulations. *Liv J Comput Mol Sci*. <https://doi.org/10.33011/livecoms.1.1.5067>
 33. Kresse G, Hafner J (1993) Ab initio molecular dynamics for liquid metals. *Phys Rev B* 47(1):558–561. <https://doi.org/10.1103/PhysRevB.47.558>
 34. Kresse G, Hafner J (1994) Ab initio molecular-dynamics simulation of the liquid–metal–amorphous–semiconductor transition in germanium. *Phys Rev B* 49(20):14251–14269. <https://doi.org/10.1103/PhysRevB.49.14251>
 35. Kresse G, Furthmüller J (1996) Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput Mater Sci* 6(1):15–50. [https://doi.org/10.1016/0927-0256\(96\)00008-0](https://doi.org/10.1016/0927-0256(96)00008-0)
 36. Kresse G, Furthmüller J (1996) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys Rev B* 54(16):11169–11186. <https://doi.org/10.1103/PhysRevB.54.11169>
 37. Hammer B, Hansen LB, Nørskov JK (1999) Improved adsorption energetics within density-functional theory using revised Perdew–Burke–Ernzerhof functionals. *Phys Rev B* 59:7413–7421. <https://doi.org/10.1103/PhysRevB.59.7413>
 38. Forster-Tonigold K, Groß A (2014) Dispersion corrected RPBE studies of liquid water. *J Chem Phys* 141:064501. <https://doi.org/10.1063/1.4892400>
 39. Monkhorst HJ (1976) Special points for Brillouin-zone integrations. *Phys Rev B* 13(12):5188–5192. <https://doi.org/10.1103/PhysRevB.13.5188>
 40. Blöchl PE (1994) Projector augmented-wave method. *Phys Rev B* 50(24):17953–17979. <https://doi.org/10.1103/PhysRevB.50.17953>
 41. Kresse G, Joubert D (1999) From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys Rev B* 59(3):1758–1775. <https://doi.org/10.1103/PhysRevB.59.1758>
 42. Vanden-Eijnden E, Ciccotti G (2006) Second-order integrators for Langevin equations with holonomic constraints. *Chem Phys Lett* 429(1):310–316. <https://doi.org/10.1016/j.cplett.2006.07.086>
 43. Andersen HC (1983) Rattle: a velocity version of the shake algorithm for molecular dynamics calculations. *J Comput Phys* 52(1):24–34. [https://doi.org/10.1016/0021-9991\(83\)90014-1](https://doi.org/10.1016/0021-9991(83)90014-1)
 44. Lankhorst D, Schriever J, Leyte JC (1982) Determination of the rotational correlation time of water by proton NMR relaxation in H₂O and some related results. *Ber Bunsenges Phys Chem* 86(3):215–221. <https://doi.org/10.1002/BBPC.19820860308>
 45. Janesko BG, Barone V, Brothers EN (2013) Accurate surface chemistry beyond the generalized gradient approximation: illustrations for graphene adatoms. *J Chem Theory Comput* 9(11):4853–4859. <https://doi.org/10.1021/ct400736w>
 46. Sakong S, Forster-Tonigold K, Groß A (2016) The structure of water at a Pt(111) electrode and the potential of zero charge studied from first principles. *J Chem Phys* 144:194701. <https://doi.org/10.1063/1.4948638>
 47. Örn Jónsson E, Dohn AO, Jónsson H (2019) Polarizable embedding with a transferable H₂O potential function I: formulation and tests on dimer. *J Chem Theory Comput* 15:6562–6577. <https://doi.org/10.1021/acs.jctc.9b00777>
 48. Dohn AO, Örn Jónsson E, Jónsson H (2019) Polarizable embedding with a transferable H₂O potential function II: application to (H₂O)_N clusters and liquid water. *J Chem Theory Comput* 15:6578–6587. <https://doi.org/10.1021/acs.jctc.9b00778>
 49. Kirchhoff B, Örn Jónsson E, Dohn AO, Jacob T, Jónsson H (2021) Elastic collision based dynamic partitioning scheme for hybrid simulations. *J Chem Theory Comput* 17:5863–5875. <https://doi.org/10.1021/acs.jctc.1c00522>
 50. Garcia-Ratés M, García-Muelas R, López N (2017) Solvation effects on methanol decomposition on Pd(111), Pt(111), and Ru(0001). *J Phys Chem C* 121(25):13803–13809. <https://doi.org/10.1021/acs.jpcc.7b05545>

51. Van den Bossche M, Skúlason E, Rose-Petruck C, Jónsson H (2019) Assessment of constant-potential implicit solvation calculations of electrochemical energy barriers for H₂ evolution on Pt. *J Phys Chem C* 123(7):4116–4124. <https://doi.org/10.1021/acs.jpcc.8b10046>
52. Basdogan Y, Maldonado AM, Keith JA (2020) Advances and challenges in modeling solvated reaction mechanisms for renewable fuels and chemicals. *Wiley Interdiscip Rev Comput Mol Sci* 10(2):e1446. <https://doi.org/10.1002/wcms.1446>
53. Basdogan Y et al (2020) Machine learning-guided approach for studying solvation environments. *J Chem Theory Comput* 16(1):633–642. <https://doi.org/10.1021/acs.jctc.9b00605>
54. Jinnouchi R, Karsai F, Kresse G (2020) Making free-energy calculations routine: combining first principles with machine learning. *Phys Rev B*. <https://doi.org/10.1103/PhysRevB.101.060201>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.