Supplementary Information for: On the challenge of obtaining an accurate solvation energy estimate in simulations of electrocatalysis

Björn Kirchhoff[1,2*], Elvar Ö. Jónsson[2], Timo Jacob[1,3,4] and Hannes Jónsson[2]

[1]Institute of Electrochemistry, Ulm University, Albert-Einstein-Allee 47, Ulm, 89081, Germany.
[2]Science Institute and Faculty of Physical Sciences, University of Iceland, VR-III, Hjarðarhagi 2, Reykjavík, 107, Iceland.
[3]Helmholtz-Institute Ulm (HIU) Electrochemical Energy Storage, Helmholtz-Straße 16, Ulm, 89081, Germany.
[4]Karlsruhe Institute of Technology (KIT), P.O. Box 3640, Karlsruhe, 76021, Germany.

*Corresponding author(s). E-mail(s): bjoern.kirchhoff@uni-ulm.de;

1 The model system

Figure S1 illustrates the 32-atomic BG sheet model with the *O, *OH, and *OOH adspecies used throughout this study. Figure S2 illustrates the BG sheet model with an *O adatom in contact with 1-4 layers of water molecules.

Interactive visualizations of the model systems can be found online at https://bjk24.gitlab.io/bg-solvation/docs/visualization.html.

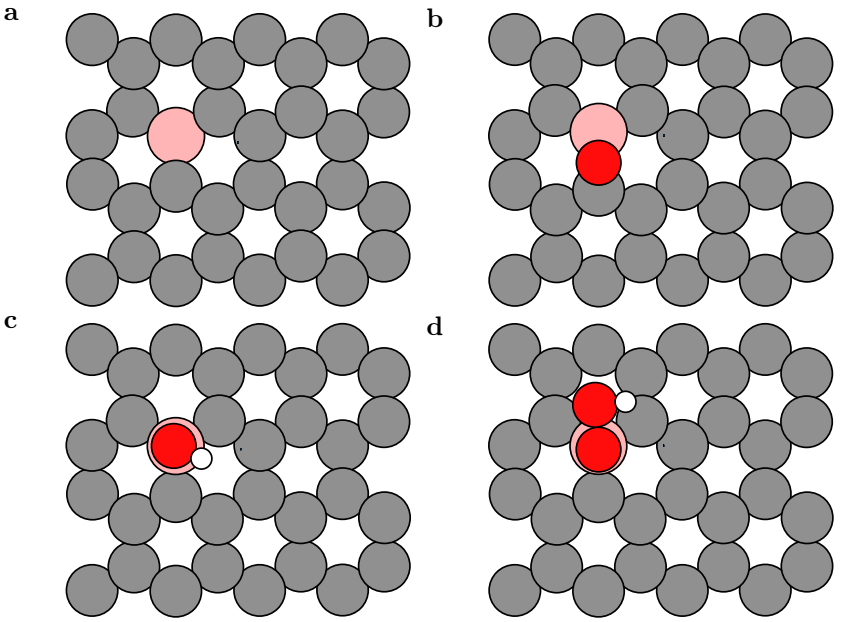

Fig. S1 Illustrations of the BG sheet model (**a**) in contact with *O (**b**), *OH (**c**), and *OOH (**d**) adspecies.

2 Density functional benchmark and ORR free energy trends

Adsorption free energy values for the ORR intermediates *O, *OH, and *OOH and, from these, thermochemical overpotentials η_{TCM} are calculated for the 32-atomic BG model using various different density functionals using the computational hydrogen electrode free energy method.[1] Potential-dependent free energy values are calculated the same way as in an earlier work on NG.[2] Using η_{TCM} as a descriptor for ORR activity, this test is performed to investigate how strongly the thermochemical results depend on the chosen density functional. In a previous study on NG, the HSE06 functional was found to

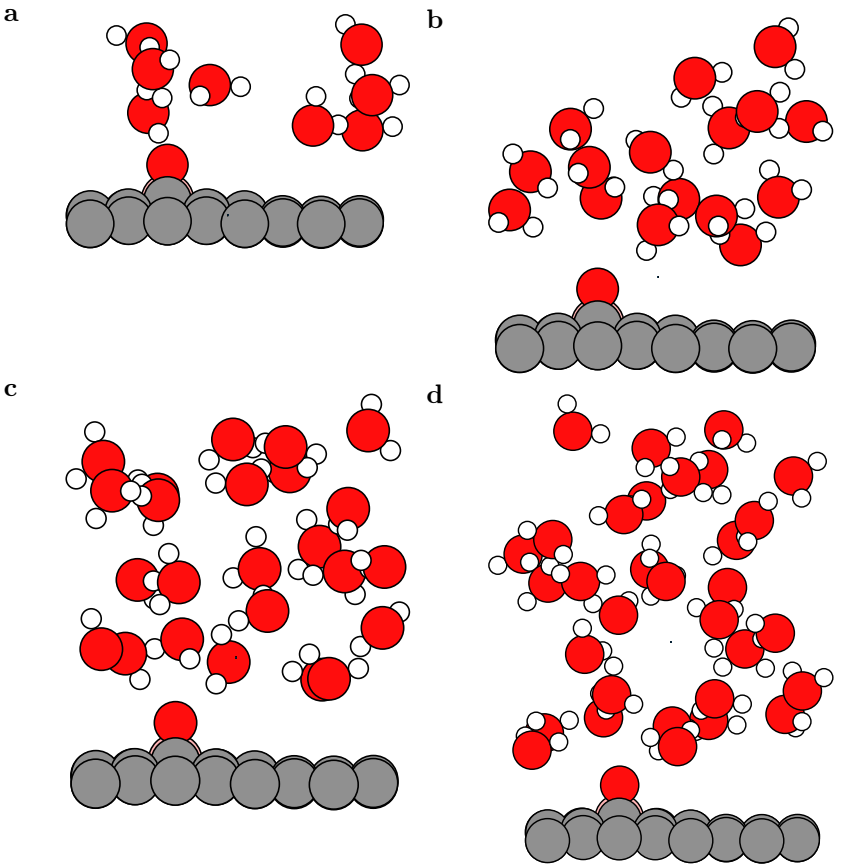

Fig. S2 Illustrations of the BG sheet model with an *O adatom in contact with 1 (**a**), 2 (**b**), 3 (**c**), and 4 layers (**d**) of water molecules.

reproduce a Diffusion Monte Carlo benchmark value most accurately out of all tested functionals.[2] Thus, the HSE06 result is used as a reference value in the following. Figure S3 shows free energy changes of the ORR intermediates on the BG model at 0 V *vs.* SHE and at the extrapolated onset potential for each functional.

The potential determining step (PDS) in all cases is the formation of the *OOH intermediate. Compared to a hypothetical ideal ORR catalyst with a free energy change of 1.23 V at each reaction step, the *OOH intermediate

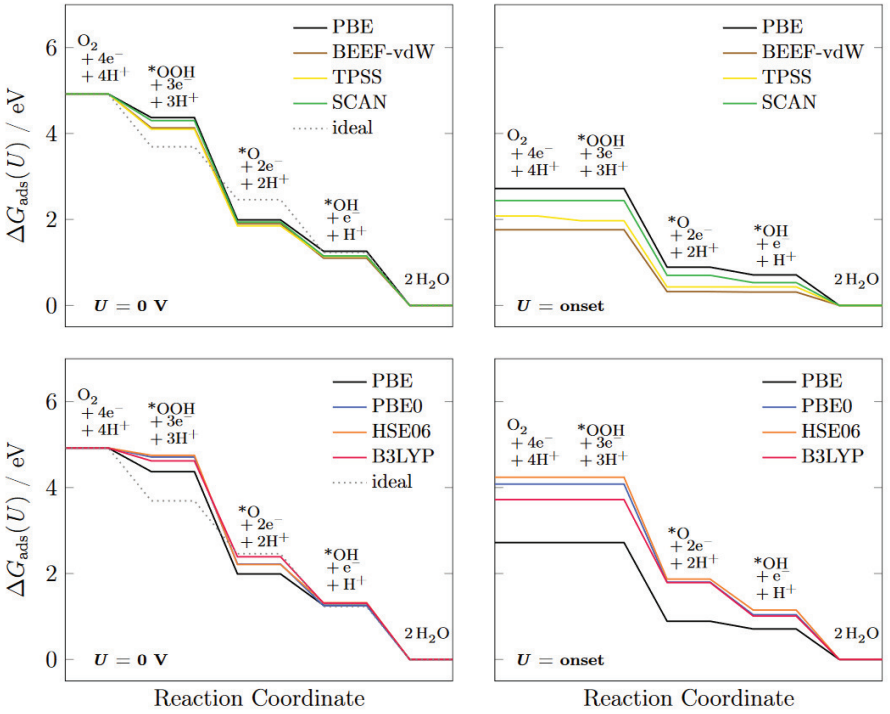

Fig. S3 Free energy diagrams for the 32-atomic BG model at 0 V obtained using GGA and meta-GGA functionals (top) as well as PBE and hybrid functionals (bottom). A hypothetical ideal ORR catalyst with a free energy change of 1.23 V at each reaction step is shown as a dotted line.

is underbound in the case of all functionals, with hybrid functionals underbinding *OOH more strongly than meta-GGA and GGA functionals. The *O intermediate is significantly overbound in the case of GGA and meta-GGA functionals which is in agreement with benchmark calculations on undoped graphene and free energy calculations on NG.[2–4] The energetic description of the *OH intermediate is similar for all tested functionals and aligns well with the hypothetical ideal catalyst. The free energy results highlight a significant issue: all functionals are in good agreement for the *OH intermediate but differ in their results for the *OOH and *O adspecies. Therefore, different functionals are bound to produce different overall trends.

To further illustrate this conclusion, thermochemical overpotentials η_{TCM} are calculated from the extrapolated onset potentials U_{onset} as

$$\eta_{\mathrm{TCM}} = 1.23\ \mathrm{V} - U_{\mathrm{onset}}. \tag{1}$$

η_{TCM} results are summarized in Table S1. The hybrid functionals HSE06 and

Table S1 Thermochemical overpotentials η_{TCM} obtained for the 32-atomic BG model with various density functionals. * Largest and smallest possible η_{TCM} obtained using standard deviations of the adsorption free energy values of the ORR intermediates, based on Bayesian error estimation using an ensemble of 2000 functionals.

Density Functional	η_{TCM} / V
HSE06	1.06
PBE0	1.02
B3LYP	0.93
PBE	0.68
SCAN	0.61
TPSS	0.52
BEEF-vdW	0.44 (0.26–0.80)*

PBE0, which constitute the most reliable result according to benchmarking,[2] perform similarly and give the highest η_{TCM} out of all tested functionals with 1.06 and 1.02 V, respectively. The B3LYP functional produces a slightly lower η_{TCM} of 0.93 V. The tested GGA and meta-GGA functionals give significantly lower η_{TCM} values of 0.68 V (PBE), 0.61 V (SCAN), 0.52 V (TPSS), and 0.44 V (BEEF-vdW). These trends are analogous to previous computational results for NG.[2] Notably, choosing a meta-GGA functional will not provide significant improvements over GGAs.

Bayesian error estimation is performed based on an ensemble of 2000 functionals generated by BEEF-vdW to obtain standard deviations for the adsorption free energy values of the ORR intermediates. Based on the error

estimation, the largest- and smallest-possible η_{TCM} value can be calculated which should be indicative of the overall uncertainty of GGA functionals for this application. The η_{TCM} range obtained this way for BEEF-vdW is 0.26–0.80 V. Notably, this range does not include values obtained by hybrid functionals. Since the HSE06 and PBE0 hybrid functionals are able to reproduce DMC benchmark values for graphene-based materials,[2–4] this result indicates that GGA functionals lack some fundamental contribution - likely exact exchange - that is necessary to accurately describe the electronic structure of this material class.

3 Convergence studies

3.1 Supercell size convergence

A supercell size convergence study is performed with the generalized gradient approximation (GGA) functional by Perdew, Burke, and Ernzerhof (PBE) as well as with the hybrid functional by Heyd, Scuseria, and Ernzerhof (HSE06) where the calculation of exact exchange was downsampled:

1. PBE-based optimization (labeled "PBE")
2. HSE06-based optimization where the k grid was reduced to the Γ point for the Hartree-Fock portion of the calculation (labeled "HSE06-fast")

The downsampling of the HSE06 functional reduces the k grid to the Γ point for the calculation of the Hartree-Fock exchange energy, thereby making minimization of the atomic coordinates computationally feasible at the hybrid DFT level.

The test below calculates energy differences according to

$$\Delta E = E_{\mathrm{tot}}^{\mathrm{ads}} - E_{\mathrm{tot}}^{\mathrm{clean}}, \tag{2}$$

where $E_{\text{tot}}^{\text{ads}}$ is the total energy of a system with an adspecies (*O, *OH, or *OOH) and $E_{\text{tot}}^{\text{clean}}$ is the total energy of the BG sheet without any adspecies. To test for supercell size convergence, it is not necessary to calculate an actual adsorption energy by taking into account the total energy of the adspecies calculated from molecules such as O_2, H_2, and H_2O because their total energy will change only in small ways as the size of the supercell increases. There is always a slight change because plane waves always fill the entire box, also for molecules, which makes the total energy dependent on the box size. However, this effect is insignificant compared to the influence of decreasing dopant and adatom concentration as a function of the supercell size. Figure S4 shows the results of this test.

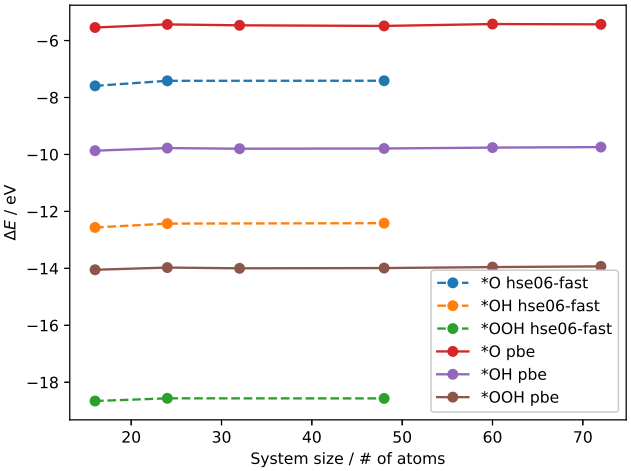

Fig. S4 Supercell size convergence study where ΔE is calculated according to equation (2) for the BG sheet in contact with O, OH, and OOH adspecies using the PBE and a downsampled HSE06 functional (see text for details).

Since these convergence tests were performed at a much earlier date than the production calculations shown in the Results section of the main manuscript, slightly different settings were used in this test compared to

the settings summarized in the Computational Details section of the main manuscript. The following list addresses the differences and their potential impact on the results:

1. This data set uses a PAW energy cutoff of ENCUT = 500. As can be seen further below in section 3.3, the choice of ENCUT = 600, which was used in the main article for geometry optimization calculations, is on the paranoid end of safe and there is no reason to assume that using a cutoff of 500 eV in this instance had an adverse effect on the results.
2. Note that the k grid was individually optimized for each system. The k grid optimization runs are not shown for the sake of brevity. The entire calculated data set is provided in a archive found under DOI:10.5281/zenodo.7684918; consult the KPOINTS files in the respective subfolders for the converged k grid settings.
3. We expect that the trends obtained from these functionals (PBE and HSE06, the latter of which constitutes a PBE functional with 25% exact exchange and a screening parameter) are fully transferable to the RPBE functional with DFT-D3 dispersion correction that was used for production calculations later on. RPBE and PBE belong to the same family of GGA-rung functionals, more specifically RPBE is a re-parameterized PBE functional optimized towards surface (adsorption) calculations. The differences between PBE and RPBE are minor, unlike for example the differences between PBE and functionals like BEEF or SCAN which use fundamentally different potential terms and would therefore require careful re-investigation.
4. The influence of DFT-D3 on the ΔE results is negligible because the adspecies are covalently bound. DFT-D3 is therefore unlikely to affect the results of this kind of convergence test but will become more important as

layers of non-covalently bound water molecules are added to the model systems. The influence of DFT-D3 on the results is investigated in more detail in the Discussion section.

This convergence test shows that the energy differences are well converged from the beginning. This result is in contrast to what was observed for NG where the the size 16 and 24 data points did not show converged results yet.[2] However, to stay consistent with the previous work on NG, we chose to use the 32 atomic model system going forward.

There is also another reason for using the slightly larger system: there is the possibility that if the system size is chosen too small, the water atoms in the individual layers become too crowded and do not have the necessary space to relax and accommodate the surface and adspecies properly. To investigate crowding effects in the lateral directions properly, the MD simulations presented later on in section Results should be repeated for a set of surface models with increasing size in x and y direction; however, such a test was not computationally feasible at the time of this study.

3.2 k grid convergence of the solvation stabilization energy

Convergence of $\Delta\Delta E_{\mathrm{solv}}$ with respect to the k grid is tested. Our hypothesis was that the k grid density of ΔE_{ads} and $\Delta\Delta E_{\mathrm{solv}}$ should be different because the interaction is fundamentally different (covalent interaction of adatom with periodic surface *vs.* non-convalent long-range interaction of solvent molecules with - mostly - the adspecies and only very lightly with the hydrophobic graphene surface). This hypothesis was strengthened by the observation that ΔE_{ads} and $\Delta\Delta E_{\mathrm{solv}}$ do not share the same dependence on the density functional.[2]

In this case $\Delta\Delta E_{\mathrm{solv}}^{\mathrm{O}*}$ for the BG system with an *O adatom is calculated as

$$\Delta\Delta E_{\mathrm{solv}}^{*\mathrm{O}} = E_{\mathrm{tot}}^{\mathrm{BG+O+solv}} - E_{\mathrm{tot}}^{\mathrm{BG+O}} - E_{\mathrm{tot}}^{\mathrm{solv}}, \tag{3}$$

where $E_{\mathrm{tot}}^{\mathrm{BG+O+solv}}$ is the total energy of the BG sheet model with the adatom and an overlayer of 8 H_2O molecules, $E_{\mathrm{tot}}^{\mathrm{BG+O}}$ is the total energy of the BG sheet model with the adatom without any solvent molecules, and $E_{\mathrm{tot}}^{\mathrm{solv}}$ is the total energy of only the 8 H_2O molecules. All total energy values are obtained as single-point results from the same starting geometry; the atomic positions are not optimized for the different subsystems. Figure S5 shows the results of this test.

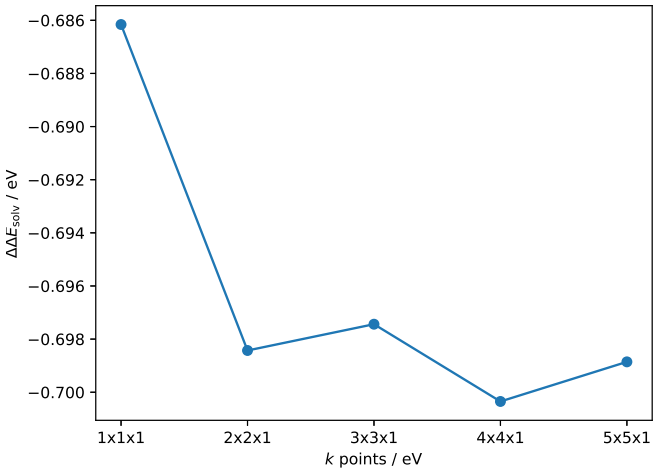

Fig. S5 Convergence study of $\Delta\Delta E_{\mathrm{solv}}$ with respect to the k grid, where $\Delta\Delta E_{\mathrm{solv}}$ is calculated according to equation (3) for the BG sheet in contact with an O adatom using the PBE functional.

From this test, $\Delta\Delta E_{\mathrm{solv}}$ results are converged using a 2x2x1 k grid. Arguably, the results can be regarded converged already at a 1x1x1 grid since the energy difference between the smallest and the next k grid is only *ca.* 0.012

eV. For the MD simulations, we erred on the side of caution and used a 3x3x1 k grid. Static calculations used various k grids depending on the exact systems; consult the data set under DOI:10.5281/zenodo.7684918 for exact k grid settings for each subset of simulations.

3.3 Convergence of the plane-augmented wave energy cutoff

The same approach as above for the supercell size was used to establish the relationship between ΔE and the plane-augmented wave energy cutoff. Only the *O adatom is tested in this case since behavior appears to be similar for all intermediates (see for example figure S4) and because the *O intermediate in particular was found to be the most notorious in benchmark calculations with NG.[2] The 32-atomic BG sheet model was used. Aside from using the PBE functional and a changing ENCUT parameter, the other simulation parameters were consistent with those summarized in the Computational Details section in the main article. Figure S6 shows the results of this test.

Results show that a PAW cutoff energy of 500 eV is sufficient to obtain converged relative energy results. For geometry optimization calculations, a cutoff of 600 eV was chosen in an attempt to err on the side of caution. However, MD simulations were performed with ENCUT = 400 due to the excessive computational cost of higher energy cutoff values. See section Discussion in the main article for an analysis of this disparity.

As above for the supercell size convergence, there is no reason to assume that this convergence test with the PBE functional would not translate to the RPBE + DFT-D3 functional combination later on as they are closely related functionals from the GGA-DFT rung.

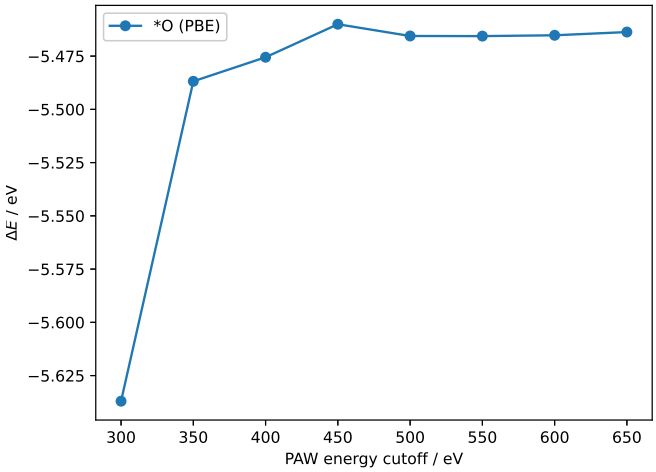

Fig. S6 Plane-augmented wave energy cutoff convergence study where ΔE is calculated according to equation (2) for the BG sheet in contact with an O adatom using the PBE functional.

4 Oxygen z distribution ($g(z)$) results

Figure S7 shows results for the z distribution of O atoms in the model systems. The distributions were obtained from the *constrained MD* and the *flexible MD* data sets. The distance pairs were calculated only from pairs that involved the surface as one of the partners; hence, the surface is located at $z = 0$ in the figures below and the distribution can be interpreted as the coordination of water atoms relative to the surface. The adspecies (*O, *OH, *OOH) are omitted from the analysis since the surface and adspecies were frozen during MD simulations and would show up as a sharp density peak higher than bands resulting from the water structure which is actually of interest.

Figure S7 shows that results from the *constrained MD* data set are smoother due to the factor 10 longer sampling. However, constraining the water molecules also appears to remove some of the fine structure observed in the case of the *flexible MD* data set; this is most apparent for the *OH and *OOH adspecies in contact with 3 layers of water molecules. This observation may

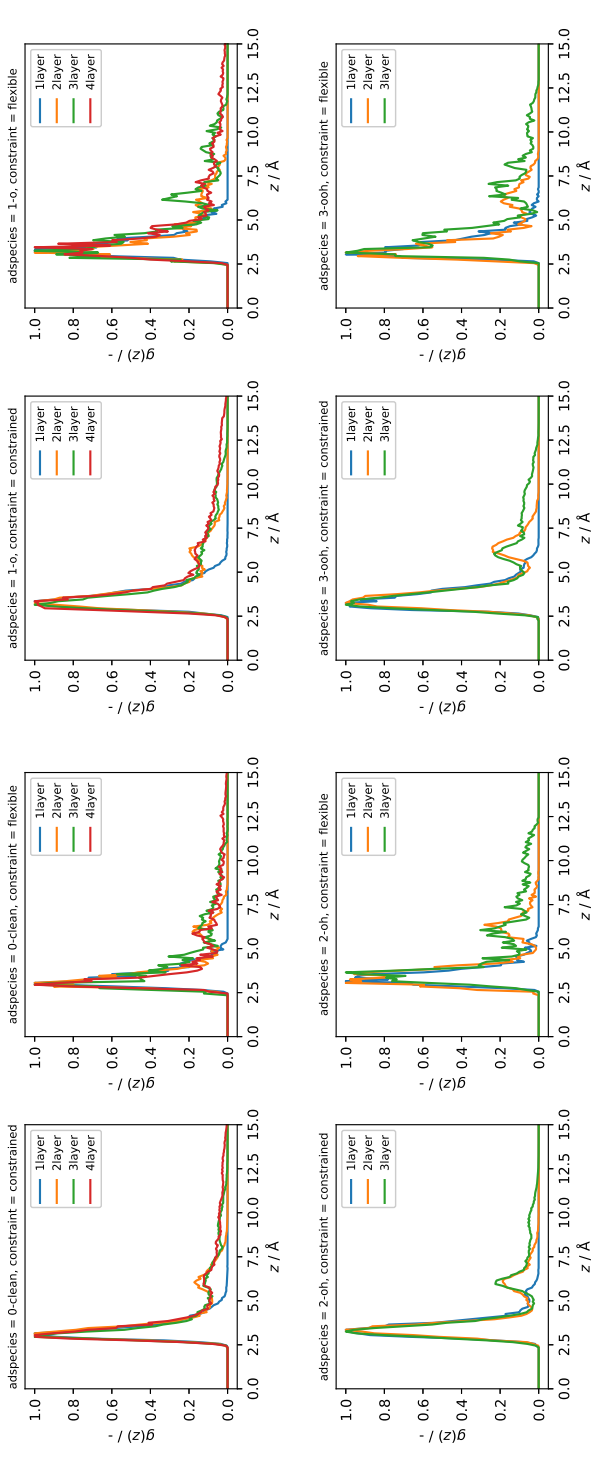

Fig. S7 Distributions of O atoms in z direction, $g(z)$, for the different systems in contact with 1–4 layers (8–32 molecules) of water: clean BG sheet model without adspecies, BG sheet with *O adatom, BG sheet with *OH admolecule, and BG sheet with *OOH admolecule. Results from the *constrained MD* simulations (100 ps total) are always shown on the left, results from the *flexible MD* (10 ps total) are shown on the right. The $g(z)$ is sampled every 100 fs in both cases. Distances are calculated between all water O atoms and the $x-y$ plane located inside the BG sheet model. Results were normalized so that the maximum $g(z)$ value in every distribution is 1.0 for better comparability.

hint towards the Rattle constraint changing the interaction with the surface and adspecies but more research is needed to make sure that this observation is not noise, *i.e.*, the result of poor sampling statistics.

5 The solvation stabilization energy obtained only from the lowest-energy MD configurations

An article by Yu *et al.* from 2011 details yet another way of obtaining $\Delta\Delta E_{\mathrm{solv}}$.[5] The group performed MD simulations with NG model systems and the ORR adspecies in contact with explicit water molecules. They then picked the lowest-energy configuration generated in each MD simulation and performed an energy minimization calculation with respect to the atomic coordinates. The minimized systems were then used to calculate solvated free energy values.

Figure S8 applies this strategy to the *flexible MD* data set.

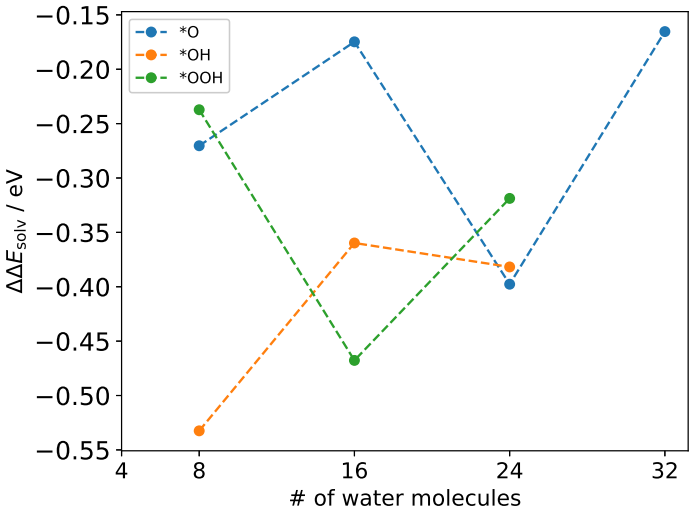

Fig. S8 $\Delta\Delta E_{\mathrm{solv}}$ obtained from only the lowest-energy configurations in the *flexible MD* data set which were subsequently energy minimized.

The closest comparison for this analysis are the results from the *resampled* data set, see figure 4 in the main article. While the trends for *OH and *OOH in figure S8 loosely resemble the trends for the *resampled* data set, $\Delta\Delta E_{\text{solv}}$ for the *O adspecies is significantly more negative than with any other analysis strategy. It can be concluded that this approach not only did not resolve the erratic trends but likely further distorted 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.

6 Influence of the MD sampling frequency on the solvation stabilization energy results

Figure S9 summarizes $\Delta\Delta E_{\text{solv}}$ results from the *flexible MD* and *constrained MD* data sets analogous to figure 4 in the main manuscript except that different intervals of sampling are tested (2 ps, 1.0 ps, 100 fs, and 10 fs represented by sampling factors 0.5, 1.0, 10.0, and 100.0, respectively). This test probes the robustness of the results against the sampling frequency and visualizes the impact that assuming more or less independent samples has on the obtained confidence intervals.

Figure S9 shows that the average $\Delta\Delta E_{\text{solv}}$ results are robust against the sampling frequency. The only significant change is observed going from sampling factor 0.5 (2 ps) to factor 1.0 (1 ps) in the case of the *flexible MD* data set. This change constitutes the the difference between 5 and 10 evaluated images for this data set. It can therefore be concluded that 10 independent images is the minimum number of images required to obtain a robust average $\Delta\Delta E_{\text{solv}}$ from this data set.

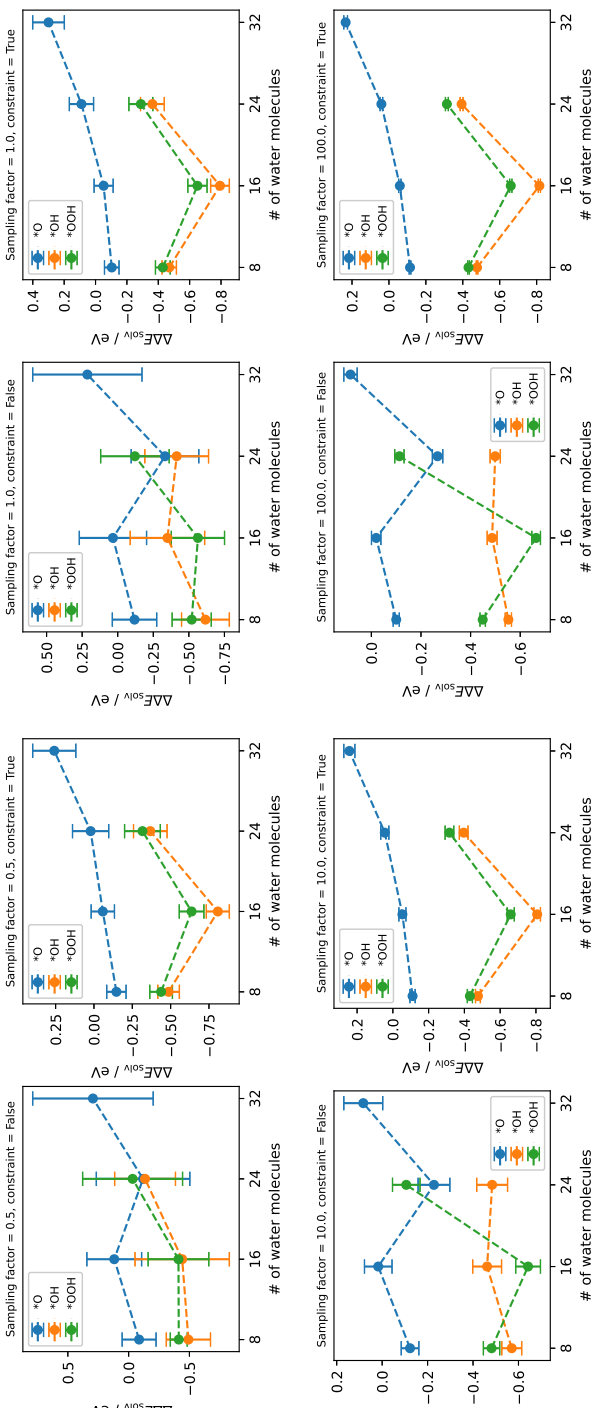

Fig. S9 Influence of the MD sampling frequency on the $\Delta\Delta E_{\mathrm{solv}}$ results. The original sampling frequency shown in the main manuscript is 1 ps, indicated in this figure as sampling factor = 1. Shown here are additional sampling factors of 0.5, 10.0, and 100.0 which correspond to sampling frequencies of 2 ps, 100 fs, and 10 fs.

The choice of sampling frequency affects the size of the error bars significantly. This result highlights that it is important to chose the sampling frequency according to physical considerations (here: correlation time of water) since only checking for convergence of the average results can create a false sense of security from oversampling the data.

7 Influence of the dipole and quadrupole correction on the results

This test is performed to check if potentially erratic dipole and quadrupole correction values, which were empirically observed in this work to sometimes occur for no apparent reason, are causing the $\Delta\Delta E_{\mathrm{solv}}$ results to be erratic. Figure S10 visualizes the dipole and quadrupole energy correction results for the *resampled* data set.

Results show that the average correction energy values are below 0.05 eV and therefore within chemical accuracy. However, the error bars for some systems are large, in particular for the *OOH adspecies in contact with 16 and 24 water molecules and the *O adspecies in contact with 24 molecules. This observation indicates that the correction can be somewhat erratic for certain systems and arrangements. Ultimately, the low absolute values for this correction are unlikely to impact results in a significant way.

8 Influence of the dispersion correction on the results

Similar to the dipole and quadrupole correction energy, the influence of the DFT-D3 dispersion correction on the obtained $\Delta\Delta E_{\mathrm{solv}}$ results is tested. Figure S11**a** shows the difference of the dispersion contributions of the clean

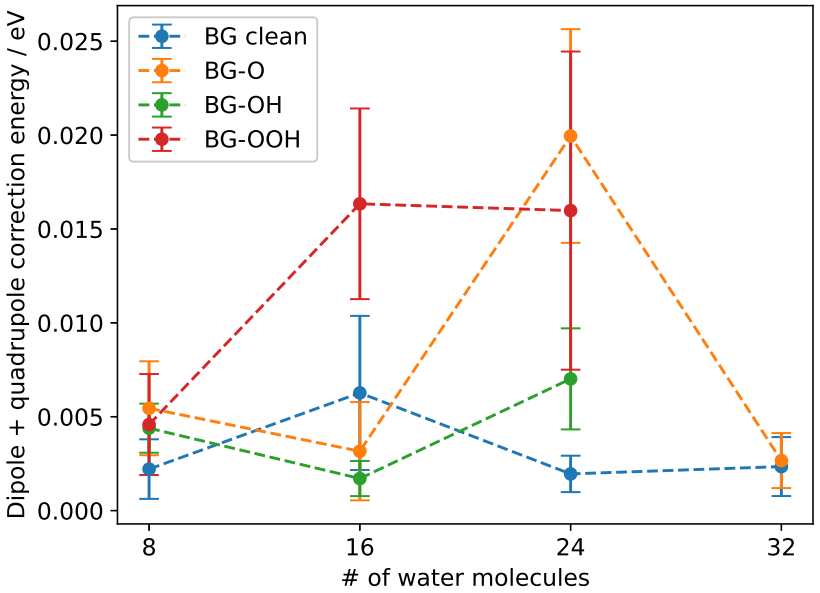

Fig. S10 Average dipole and quadrupole energy correction values for the systems in the *resampled* data set. While large fluctuations are noted in particular for the *OOH adspecies in contact with 24 and 16 water molecules and the *O adspecies in contact with 24 water molecules, the fluctuations remain below the limit of chemical accuracy (< 0.05 eV). The error bars indicate the two-sided 95 % confidence interval.

BG surface and the BG surface with an adspecies (X = O, OH, OOH):

$$\Delta E_{\text{disp}} = E_{\text{disp}}^{\text{BG-X}} - E_{\text{disp}}^{\text{BG}}. \tag{4}$$

Figure S11**b** reproduces the $\Delta\Delta E_{\text{solv}}$ results for the *resampled* data set shown in Figure 4 in the main manuscript but with the dispersion energy contribution removed from the total energy values before calculating $\Delta\Delta E_{\text{solv}}$.

Results indicate that while the dispersion correction is significant in terms of absolute values, removing the E_{disp} does not stabilize the $\Delta\Delta E_{\text{solv}}$ trends either (figure S11**b**). There is, however, an important caveat to this test: dispersion correction was included during the MD simulations from which the *resampled* data set was generated and also during minimization calculations

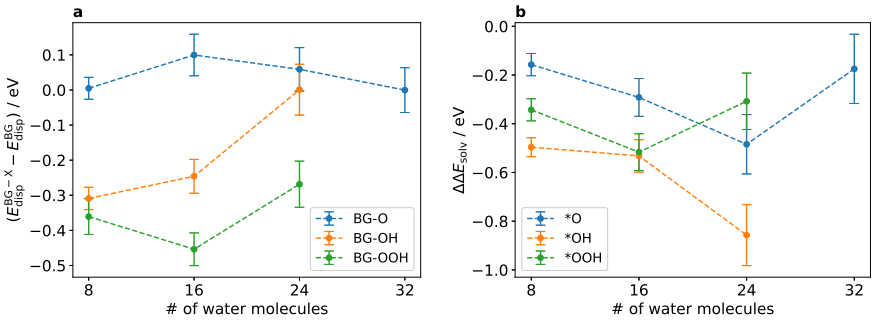

Fig. S11 Influence of the dispersion correction on the calculated $\Delta\Delta E_{\mathrm{solv}}$ results. **a** Absolute dispersion correction energy for each tested system averaged over the data set of 20 images from the *resampled* data set as outlined in the main article. Shown here is $\Delta E_{\mathrm{disp}} = E_{\mathrm{disp}}^{\mathrm{BG-X}} - E_{\mathrm{disp}}^{\mathrm{BG}}$, *i.e.*, the difference of the dispersion contributions of the clean BG surface and the BG surface with an adspecies (X = O, OH, OOH). **b** $\Delta\Delta E_{\mathrm{solv}}$ trends for the *resampled* data set where E_{disp} has been deducted from the total energy values before calculating $\Delta\Delta E_{\mathrm{solv}}$. The error bars indicate the two-sided 95 % confidence interval.

of the structures in the *resampled* data set. Hence, this *a posteriori* removal of the dispersion contribution can only be a rough indicator of its influence. The data sets would need to be reproduced completely without dispersion correction to conclusively rule out this parameter.

9 Influence of the simulation box dimensions

Table S2 summarizes the total energy values of the reference systems without water molecules from the *resampled* data set used to calculate $\Delta\Delta E_{\mathrm{solv}}$. The reference systems were generated from the parent systems that include water molecules by removing the latter. This step was necessary because the simulation box dimensions are different depending on how many water molecules are included. This table illustrates the differences introduced into the total energy by variable cell dimensions. The differences are < 0.01 eV in all cases and, even if the cell dimension had not been corrected for in this way, are unlikely to distort the simulation results in a meaningful way.

Table S2 Summary of total energy values of the reference systems without water molecules used to calculate $\Delta\Delta E_{\mathrm{solv}}$. The reference systems are from the *resampled* data set.

System	# of H_2O	E_{tot} / eV
BG sheet (clean)	8	-290.425863
	16	-290.425863
	24	-290.429015
	32	-290.426769
BG-O*	8	-295.994912
	16	-295.994912
	24	-295.998699
	32	-295.996450
BG-OH*	8	-300.638397
	16	-300.638397
	24	-300.638397
BG-OOH*	8	-304.701104
	8	-304.701104
	8	-304.701104

10 Influence of energy-minimizing the non-solvated reference systems

Another potential source of inconsistency is the way that total energy values for the non-solvated reference systems are obtained. In the main article, the configurations of the non-solvated systems were obtained by removing the water molecules from the solvated systems and minimizing the resulting atomic configurations. However, with this approach, $\Delta\Delta E_{\mathrm{solv}}$ does not only contain the interaction between the surface-adspecies system with the water molecules but also the rearrangement energy from the relaxation of the reference system. Figure S12 explores for the *one-shot minimization* data set whether *not* minimizing the reference systems, *i.e.*, obtaining the reference energy values from single-point calculations on the formerly-solvated systems where water molecules have been removed, makes a difference.

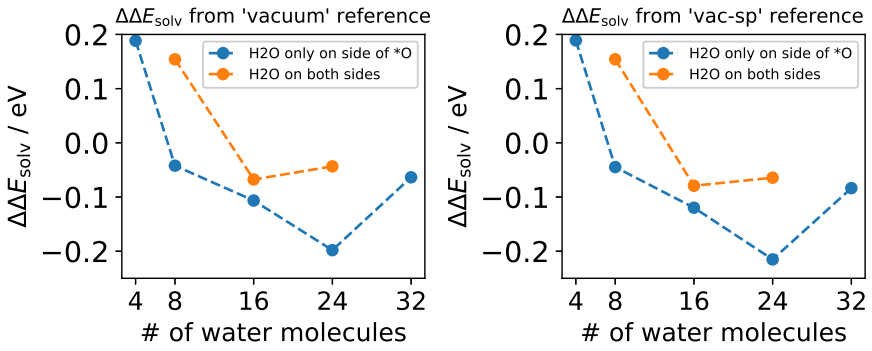

Fig. S12 Exploring the influence of energy-minimizing the non-solvated reference systems. Left: non-solvated reference configurations were energy-minimized with respect to the atomic coordinates ('vacuum'). Right: non-solvated reference systems were not energy-minimized and energy values were obtained from single-point calculations ('vac-sp').

Results show that the differences are minimal at best. The relative trend does not change at all but absolute values are slightly more negative in the case of the single-point reference calculations compared to the minimized reference calculations for 16, 24, and 32 water molecules.

11 Influence of freezing the BG sheet

The *flexible MD* simulation of BG-OOH in contact with 8 (flexible) water molecules was accidentally performed with a non-constrained BG sheet. While the data shown in the main article was obtained with the correctly constrained model, this mistake allows to probe the influence of this geometry constraint. Figure S13 compares the E_{tot} *vs.* t and T *vs.* t trends of *flexible MD* simulations with the unconstrained BG sheet (**a**) and the properly constrained BG

sheet (**b**). Figure S14 compares the surface-O $g(z)$ distributions obtained from *flexible MD* simulations with these two models.

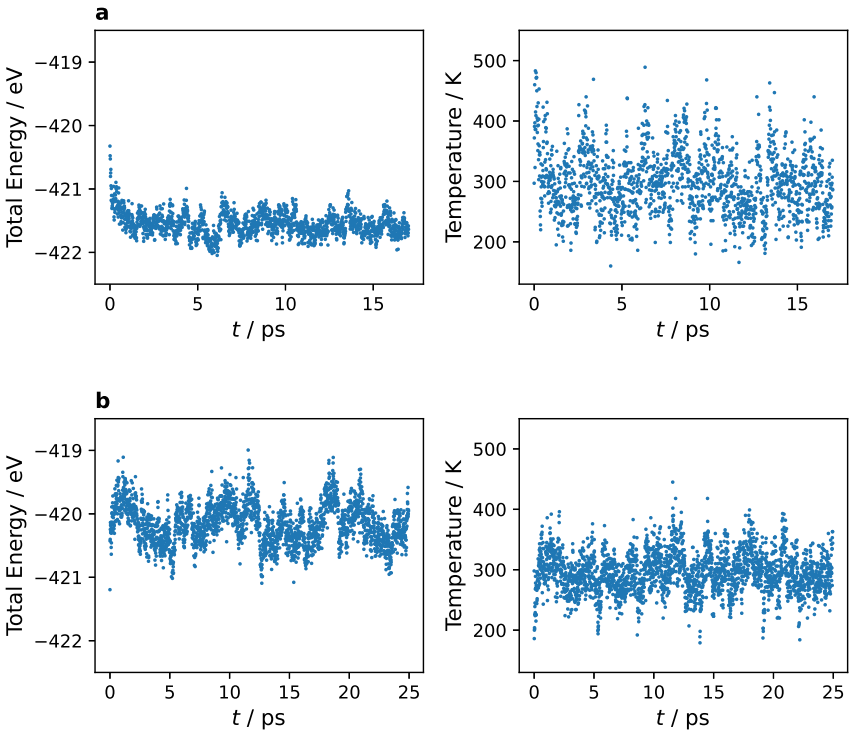

Fig. S13 Comparison of E_{tot} *vs.* t and T *vs.* t trends of *flexible MD* simulations with the properly constrained BG sheet (**a**) and the unconstrained BG sheet (**b**).

Figure S13 shows that the total energy and temperature fluctuations are significantly increased if the BG sheet with the adspecies is not constrained (**b**). This result may indicate that significantly more sampling would be required in the case of the non-constrained BG sheet to obtain a good estimate of $\Delta\Delta E_{\text{solv}}$ with a small-enough error bar. Figure S14 does not show any significant differences between the two systems, indicating that freezing the BG sheet and adspecies does not significantly change the interaction with the first layer of solvent molecules.

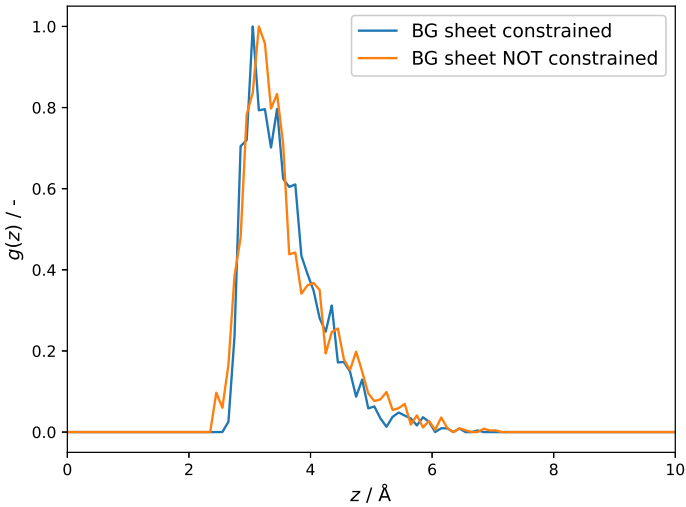

Fig. S14 Comparison surface-O $g(z)$ distributions obtained from *flexible MD* simulations using the constrained and non-constrained BG sheet.

References

[1] Nørskov, J. K. *et al.* Origin of the overpotential for oxygen reduction at a fuel-cell cathode. *J. Phys. Chem. B* **108** (46), 17886–17892 (2004). https://doi.org/10.1021/jp047349j .

[2] Kirchhoff, B. *et al.* Assessment of the accuracy of density functionals for calculating oxygen reduction reaction on nitrogen-doped graphene. *J. Chem. Theory Comput.* **17**, 6405–6415 (2021). https://doi.org/10.1021/acs.jctc.1c00377 .

[3] Hsing, C. R., Wei, C. M. & Chou, M. Y. Quantum monte carlo investigations of adsorption energetics on graphene. *J. Phys. Condens. Matter* **24** (39), 395002 (2012). https://doi.org/10.1088/0953-8984/24/39/395002 .

[4] Janesko, B. G., Barone, V. & Brothers, E. N. Accurate surface chemistry beyond the generalized gradient approximation: Illustrations for graphene adatoms. *J. Chem. Theory Comput.* **9** (11), 4853–4859 (2013). https://doi.org/10.1021/ct400736w .

[5] Yu, L., Pan, X., Cao, X., Hu, P. & Bao, X. Oxygen reduction reaction mechanism on nitrogen-doped graphene: A density functional theory study. *J. Catal.* **282**, 183–190 (2011). https://doi.org/10.1016/J.JCAT.2011.06.015 .