

Supplementary Material: *Ab Initio* Surface Chemistry with Chemical Accuracy: Application to Water on Metal Oxides

Hong-Zhou Ye^{*,†} and Timothy C. Berkelbach^{*,†,‡}

[†]*Department of Chemistry, Columbia University, New York, New York 10027, USA*

[‡]*Initiative for Computational Catalysis, Flatiron Institute, New York, NY 10010, USA*

E-mail: hzyechem@gmail.com; t.berkelbach@columbia.edu

Note: figures and equations appearing in the main text will be referred to as “Fig. Mxxx” and “Eq. Mxxx” in this Supplementary Material document.

Contents

S1 Overall computational details	4
S1.1 Geometries	4
S1.2 SCAN+D3, PBE+ <i>U</i> +D3, PBE0+D3, HF, and MP2	5
S1.3 LNO-CCSD(T)	6
S1.4 Adsorption energy calculation	11
S1.5 Vibrational corrections	12
S2 Water on the α-Al₂O₃(0001) surface	13
S2.1 Atomic structure	13
S2.2 Converged energetics	13
S2.3 Slab size, surface size, basis size, and Brillouin zone sampling	14
S2.4 Infinite vacuum correction	18
S2.5 Vibrational corrections	19
S3 Water on the rutile TiO₂(110) surface	21
S3.1 Atomic structure	21
S3.2 Converged energetics	21
S3.3 Slab size, surface size, basis size, and Brillouin zone sampling	22
S3.4 Infinite vacuum correction	26
S3.5 Vibrational corrections	27
S4 Optimized geometries	29
S4.1 Bulk lattice constants	29
S4.2 Optimized geometries	29
S4.3 Energy uncertainty from the geometry	29
S5 Timing	30

S1 Overall computational details

S1.1 Geometries

All geometries are optimized at the PBE(+D3) level of theory using the Quantum Espresso (QE) v6.4 code^{1,2} with an energy convergence threshold of 0.2 meV/cell and a force convergence threshold of 0.05 eV/Å. The projector augmented wave pseudopotentials for PBE as obtained from the QE official site are used for all elements (Table S1). For both surface systems (Al₂O₃ and TiO₂), bulk geometries are first optimized, from which clean surfaces are constructed and optimized. The molecularly and dissociatively adsorbed water-surface geometries are then constructed and optimized, followed by climbing image nudged elastic band (CI-NEB) calculations to find the transition-state (TS) geometries. The D3 dispersion correction by Grimme and co-workers³ is used for the PBE geometry optimization of both clean surfaces and surfaces with adsorbates but not for bulk, where we found the effect of the dispersion correction to be small (see Section S4.1). Other parameters affecting the geometry relaxation, including the plane-wave kinetic energy cut-off, the k -point mesh for Brillouin zone sampling, and the size of the slab model are detailed in Section S2 for the Al₂O₃ surface and Section S3 for the TiO₂ surface, respectively. The final optimized geometries are provided in Section S4.2. We also estimate the effect of the employed PBE+D3 geometries on the calculated adsorption and reaction energetics in Section S4.3.

Table S1: The pseudopotential file (as obtained from the Quantum Espresso official site) and valence electrons for each element used by the plane-wave DFT calculations in this work.

Element	PP file	Valence electrons
Al	Al.pbe-n-kjpaw_psl.1.0.0.UPF	$3s^23p^1$
Ti	Ti.pbe-spn-kjpaw_psl.1.0.0.UPF	$3s^23p^63d^2$
O	O.pbe-n-kjpaw_psl.1.0.0.UPF	$2s^22p^4$
H	H.pbe-kjpaw_psl.1.0.0.UPF	$1s^1$

S1.2 SCAN+D3, PBE+*U*+D3, PBE0+D3, HF, and MP2

Single-point calculations are performed at the PBE+D3-optimized geometries from above using different levels of theory to obtain the adsorption energy, reaction energies, and barrier heights reported in the main text. These single-point calculations were performed using a combination of QE and the PySCF package.⁴⁻⁶ All numbers are carefully converged to within 1 kcal/mol with respect to the basis set size, the surface size, the slab size, and Brillouin zone sampling, as detailed in Sections S2 and S3 for the two surface systems, respectively. The D3 dispersion correction is included in all reported DFT energies (PBE, SCAN, PBE0, and PBE+*U*). For PBE and PBE+*U*, we use the D3 energy for PBE calculated from QE. For SCAN (whose D3 parameters are not available in QE) and PBE0 (for which we performed calculations with PySCF), we calculate the respective D3 energies using the `simple-dftd3` code (<https://github.com/dftd3/simple-dftd3>), and we verified that this code gives the same D3 energy for PBE compared to QE.

The PBE+*U* calculations are performed using QE with the simplified version of Cococcioni and de Gironcoli⁷ (`lda_plus_u_kind = 0`) and the atomic projector (`U_projection_type = 'atomic'`). The SCAN calculations are performed using QE with the norm-conserving Hartwigsen-Goedecker-Hutter (HGH) pseudopotentials^{8,9} optimized for PBE. All other calculations, including the wavefunction calculations and the hybrid DFT calculations with the PBE0 functional, are performed using the PySCF code⁴⁻⁶ with translational symmetry-adapted Gaussian basis functions. The Goedecker-Teter-Hutter (GTH) pseudopotentials optimized for HF¹⁰ or PBE^{8,9} (for PBE0 only) are employed, with the same number of valence electrons as listed in Table S1. The correlation-consistent Gaussian basis set series optimized for the GTH pseudopotentials from Ref. 11, GTH-cc-pVXZ with $X = D, T, Q$ (henceforth denoted by XZ), are used for all elements except for Ti, for which we performed similar basis optimization as in Ref. 11 to generate the corresponding GTH-cc-pVXZ series. We verified at PBE level that the reaction energetics calculated from PySCF agree within 1 kcal/mol with those from QE when using a TZ basis set. The four-center electron-repulsion integrals are handled using the fast periodic density fitting algorithm from Ref. 12–14. The fitting basis sets are optimized for the GTH-cc-pVXZ series of these elements by

minimizing the error of HF and MP2 energies of molecular monohydrides. We verified that the error in the calculated reaction energetics (from both mean-field and correlated calculations) introduced by these optimized fitting basis sets are less than 0.1 kcal/mol by comparing to a large even-tempered fitting basis set generated by PySCF. All basis set data can be found in https://github.com/hongzhouye/supporting_data/tree/main/2023/arXiv%3A2309.14640.

For HF and PBE0 calculations, the finite-size error of the exact, nonlocal exchange energy arising from the integrable divergence at $\mathbf{G} = \mathbf{0}$ is treated using a Madelung constant correction^{15–17}. This yields an energy with an asymptotic $1/N_k$ convergence to the thermodynamic limit (TDL), with N_k the number of k -points used to sample the Brillouin zone. These “finite-size error-corrected” orbital energies are used to obtain both the MP2 correlation energy and the (T) part of the CCSD(T) correlation energy. In contrast, CCSD calculations are independent of the use of Madelung constant corrections, as long as they are consistently applied.

S1.3 LNO-CCSD(T)

LNO-CCSD(T) calculations are performed using an in-house implementation in a developer version of PySCF based on the molecular LNO-CCSD(T) theory by Kállay and co-workers.^{18–20} In PySCF, most methods implemented for molecular calculations are automatically applicable to solids with Γ -point Brillouin zone sampling by replacing the preceding molecular mean-field calculations (HF in our case) with the corresponding periodic version. For the general case where the Brillouin zone is sampled by an evenly spaced, Γ -included k -point mesh, we first transform the k -point HF orbitals in the original simulation cell into the corresponding Γ -point HF orbitals in the BvK supercell and then perform LNO-CCSD(T) calculations therein.

For completeness, we briefly review the LNO-CCSD(T) theory here and guide the reader to the original publications^{18–20} for details. All systems studied in this work are closed-shell and described by a spin-restricted HF reference determinant $|\Phi_0\rangle$, with canonical HF orbitals ψ_p , orbital energies ϵ_p , and total energy E_{HF} . We use i, j, k for N_o occupied orbitals, a, b, c for N_v virtual orbitals, and p, q, r, s for N unspecified molecular orbitals. In this basis, the electronic Hamiltonian

is

$$H = \sum_{pq,\sigma}^N h_{pq} a_{p\sigma}^\dagger a_{q\sigma} + \frac{1}{2} \sum_{pqrs,\sigma\sigma'}^N V_{pqrs} a_{p\sigma}^\dagger a_{q\sigma'}^\dagger a_{s\sigma'} a_{r\sigma} \quad (\text{S1})$$

with $V_{pqrs} = (pq|rs)$ in (11|22) notation.

We choose the generalized Pipek-Mezey method²¹ with a minao projector to localize the HF occupied orbitals

$$\phi_I = \sum_i^{N_o} U_{iI} \psi_i \quad (\text{S2})$$

Representative localized occupied orbitals are shown in Fig. M1C for water-Al₂O₃ and M3B for water-TiO₂. For each localized occupied orbital, ϕ_I , one constructs a local active space \mathcal{P}_I by augmenting ϕ_I with selected LNOs from second-order Møller-Plesset perturbation theory²² (MP2). Specifically, one computes the occupied-occupied and the virtual-virtual blocks of the MP2 density matrix

$$D_{ij}^I = \sum_{ab} t_{iaIb}^{(1)} [2t_{jaIb}^{(1)} - t_{IaIb}^{(1)}] \quad (\text{S3})$$

$$D_{ab}^I = \sum_{jc} t_{IaIc}^{(1)} [2t_{IbIc}^{(1)} - t_{IbIc}^{(1)}] \quad (\text{S4})$$

where

$$t_{IaIb}^{(1)} = \frac{(Ia|Ib)}{\tilde{\epsilon}_I + \epsilon_j - \epsilon_a - \epsilon_b} \quad (\text{S5})$$

is an approximate MP2 amplitude with $\tilde{\epsilon}_I = \langle \phi_I | f | \phi_I \rangle$, and f is the Fock operator. Diagonalizing the virtual-virtual block

$$D_{ab}^I = \sum_c \xi_c^I X_{ac}^I X_{bc}^I, \quad (\text{S6})$$

gives the virtual LNOs associated with ϕ_I , i.e., $\phi_b = \sum_c X_{cb}^I \psi_c$. For the occupied LNOs, we follow Ref. 18 and diagonalize

$$\tilde{D}_{ij}^I = \sum_{kl} Q_{ik}^I D_{kl}^I Q_{lj}^I = \sum_k^{N_o-1} \xi_k^I X_{ik}^I X_{jk}^I \quad (\text{S7})$$

where $Q_{ij}^I = \delta_{ij} - U_{iI} U_{jI}$ projects out ϕ_I from D_{ij}^I to prevent it from mixing with other occupied

orbitals, giving the occupied LNOs $\phi_j = \sum_k X_{kj}^I \psi_k$. The eigenvalues ξ_p^I , which are between 2 and 0, quantify the importance of a given LNO to the electron correlation of localized orbital ϕ_I . In practice, we construct the local active space \mathcal{P}_I by keeping those LNOs satisfying

$$\xi_i \geq \lambda_o, \quad \xi_a \geq \lambda_v, \quad (\text{S8})$$

for some user-selected thresholds λ_o and λ_v , producing $n = n_o + n_v$ orbitals. Typically, n is much less than N and, for gapped systems such as the surface systems studied in this work, does not increase with system size for a targeted level of accuracy. In Fig. M1C and M3B, we also show the density of the virtual LNOs generated for the representative localized occupied orbitals with $\lambda_v = 10^{-5}$, 10^{-6} , and 10^{-7} .

A local Hamiltonian is then constructed by projecting H into \mathcal{P}_I

$$H_I = \sum_{pq \in \mathcal{P}_I, \sigma}^n f_{pq}^I a_{p\sigma}^\dagger a_{q\sigma} + \frac{1}{2} \sum_{pqrs \in \mathcal{P}_I, \sigma\sigma'}^n V_{pqrs} a_{p\sigma}^\dagger a_{q\sigma'}^\dagger a_{s\sigma'} a_{r\sigma} \quad (\text{S9})$$

where

$$f_{pq}^I = h_{pq} + \sum_{j \notin \mathcal{P}_I}^{N_o - n_o} (2V_{pqjj} - V_{pjjq}), \quad (\text{S10})$$

which includes a frozen core contribution. Solving the CCSD amplitude equations with H_I gives the local CCSD amplitudes t_{ia} and t_{iajb} in \mathcal{P}_I , from which the local contribution of the CCSD(T) correlation energy $E^{(I)}$ can be evaluated by

$$E_{\text{CCSD(T)}}^{(I)} = \sum_{kl \in \mathcal{P}_I} U_{kl} [G_{kl}^{(I), \text{CCSD}} + G_{kl}^{(I), \text{(T)}}] U_{ll} \quad (\text{S11})$$

where

$$G_{kl}^{(I), \text{CCSD}} = \sum_{iab \in \mathcal{P}_I} (t_{iakb} + t_{ia} t_{kb}) (2V_{ialb} - V_{ibla}) \quad (\text{S12})$$

and

$$\begin{aligned}
G_{kl}^{(I),(T)} = & -\frac{1}{3} \sum_{a \geq b \geq c \in \mathcal{P}_I} f_{abc} \sum_{ij \in \mathcal{P}_I} \left\{ \right. \\
& w_{ijk}^{abc} \left(+8v_{ijl}^{abc} - 5v_{ikj'}^{abc} - 2v_{jil}^{abc} + 2v_{jli}^{abc} + 2v_{lij}^{abc} - 5v_{lji}^{abc} \right) + \\
& w_{ikj}^{abc} \left(-5v_{ijl}^{abc} + 8v_{ilj}^{abc} + 2v_{jil}^{abc} - 2v_{jli}^{abc} - 5v_{lij}^{abc} + 2v_{lji}^{abc} \right) + \\
& \left. w_{kij}^{abc} \left(+2v_{ijl}^{abc} - 5v_{ilj}^{abc} - 5v_{jil}^{abc} + 2v_{jli}^{abc} + 8v_{lij}^{abc} - 2v_{lji}^{abc} \right) \right\}
\end{aligned} \tag{S13}$$

with

$$f_{abc} = \begin{cases} \frac{1}{6}, & a = c, \\ \frac{1}{2}, & a = b \text{ or } b = c, \\ 1, & \text{otherwise,} \end{cases} \tag{S14}$$

$$w_{ijk}^{abc} = W_{ijk}^{abc} / \sqrt{-D_{ijk}^{abc}}, \quad v_{ijk}^{abc} = V_{ijk}^{abc} / \sqrt{-D_{ijk}^{abc}} \tag{S15}$$

$$W_{ijk}^{abc} = P_{ijk}^{abc} \left(\sum_d V_{bdai} t_{kj}^{cd} - \sum_l V_{ckjl} t_{il}^{ab} \right) \tag{S16}$$

$$V_{ijk}^{abc} = W_{ijk}^{abc} + \frac{1}{2} P_{ijk}^{abc} (V_{aibj} t_k^c) \tag{S17}$$

$$D_{ijk}^{abc} = \epsilon_{ii} + \epsilon_{jj} + \epsilon_{kk} - \epsilon_{aa} - \epsilon_{bb} - \epsilon_{cc} \tag{S18}$$

where $P_{ijk}^{abc}(\cdot)$ generates 6-fold permutation

$$P_{ijk}^{abc}(X_{ijk}^{abc}) = X_{ijk}^{abc} + X_{ikj}^{acb} + X_{jik}^{bac} + X_{jki}^{bca} + X_{kij}^{cab} + X_{kji}^{cba} \tag{S19}$$

The remaining correlation contribution from orbitals $\notin \mathcal{P}_I$ can be included at $O(N^5)$ cost by invoking a composite correction at MP2 level, leading to the final expression for the LNO-CCSD(T) total correlation energy

$$E_c = \sum_I^{N_o} \left(E_{\text{CCSD(T)}}^{(I)} - E_{\text{MP2}}^{(I)} \right) + E_{\text{MP2,full}} \tag{S20}$$

where $E_{\text{MP2}}^{(I)}$ is the MP2 correlation energy evaluated in \mathcal{P}_I with H_I and $E_{\text{MP2,full}}$ is the MP2 correla-

tion energy for the entire system. The latter can also be decomposed into contributions from each localized orbital

$$E_{\text{MP2,full}} = \sum_I \sum_{jab} \tilde{t}_{Iajb}^{(1)} (2V_{Iajb} - V_{Ibja}) = \sum_I E_{\text{MP2,full}}^{(I)} \quad (\text{S21})$$

Note that the summation over jab in Eq. (S21) is not restricted to \mathcal{P}_I and \tilde{t}_{Iajb} is the exact MP2 amplitude for the entire system with one occupied index being transformed into ϕ_I , which is not the same as the approximate local MP2 amplitudes defined in Eq. (S5). Combining Eqs. (S20) and (S21), we obtain the final expression of the LNO-CCSD(T) correlation energy

$$E_c = \sum_I E_{\text{CCSD(T)}}^{(I)} - E_{\text{MP2}}^{(I)} + E_{\text{MP2,full}}^{(I)} = \sum_I E^{(I)}. \quad (\text{S22})$$

For the purpose of calculating reaction energetics as in this work, the local nature of a chemical reaction allows us to further reduce the computational cost by invoking a strategy resembling the spirit of many fragment embedding methods. Specifically, we can choose to use CCSD(T) to evaluate $E^{(I)}$ for only a subset of $M_o \leq N_o$ localized occupied orbitals, with the contributions from the remaining $N_o - M_o$ localized occupied orbitals being accounted for by the MP2 composite correction, leading to a modified energy expression from Eq. (S20)

$$E_c = \sum_I^{M_o} (E_{\text{CCSD(T)}}^{(I)} - E_{\text{MP2}}^{(I)}) + E_{\text{MP2,full}}. \quad (\text{S23})$$

The final reaction energetics calculated from Eq. (S23) thus need to be converged with respect to both the LNO truncation parameters defined in Eq. (S8) and the embedded cluster size M_o . We detail the convergence of reaction energetics for the two systems studied in this work in Sections S2.3 and S3.3, respectively.

Summarizing the steps and cost of an LNO calculation, there are three parts.

1. Full-system MP2, which is required by both the LNO construction [Eqs. (S3)–(S7)] and the MP2 composite correction [Eq. (S20)] and scales as $O(N^5)$.
2. N_o independent integral transformations, which are required by the local Hamiltonian con-

struction [Eq. (S9)] and scale as $O(N^4n)$ each, but embarrassingly parallel in N_o .

3. Independent correlated calculations of all local Hamiltonians, which scale as N_o times the cost of a calculation at the desired level of theory in the local active space, i.e., n^6 for CCSD and n^7 for CCSD(T).

For the LNO-CCSD(T) calculations generating the final numbers reported in the main text, solving the local Hamiltonians consists of 90% of the CPU time and hence dominates the computational cost. This is due to the relatively large number of LNOs kept in the subspace for fully converging the calculated energetics to an accuracy well within 1 kcal/mol, as discussed in the main text. For calculations of an even larger system or with a looser LNO truncation threshold, the $O(N^5)$ steps may dominate the computational cost. Although not explored in this work, many numerical techniques such as local domain-based approximations¹⁹ and Laplace transform methods²⁰ have been exploited to make these steps linear scaling for molecular LNO-CCSD(T) calculations. Such advances can also be extended to the periodic LNO-CCSD(T) described here.

S1.4 Adsorption energy calculation

The adsorption energies from PBE0+D3, MP2 and LNO-CCSD(T) are calculated as follows in PySCF

$$E_{\text{ads}} = E_{\text{int}} + E_{\text{surf-relax}} + E_{\text{mol-relax}} \quad (\text{S24})$$

where

$$E_{\text{int}} = E(\text{surf} + \text{mol}) - E(\text{surf} + \text{ghost-mol}) - E(\text{ghost-surf} + \text{mol}) \quad (\text{S25})$$

is the adiabatic interaction energy corrected for the basis set superposition error,

$$E_{\text{surf-relax}} = E(\text{surf}) - E[\text{surf}(\text{relaxed})] \quad (\text{S26})$$

is the surface relaxation energy, and

$$E_{\text{mol-relax}} = E(\text{mol}) - E[\text{mol}(\text{relaxed})] \quad (\text{S27})$$

is the molecular relaxation energy. Both E_{int} and $E_{\text{surf-relax}}$ are evaluated with periodic boundary conditions, while $E_{\text{mol-relax}}$ is evaluated using the molecular code with the same GTH pseudopotential used for the periodic calculations.

S1.5 Vibrational corrections

We evaluate the vibrational zero-point energy

$$\text{ZPE} = \sum_{\alpha} \frac{1}{2} \hbar \omega_{\alpha} \quad (\text{S28})$$

and the temperature-dependent vibrational free energy

$$A_{\text{vib}}(T) = -RT \sum_{\alpha} \ln \frac{e^{\beta \hbar \omega_{\alpha}/2}}{e^{\beta \hbar \omega_{\alpha}} - 1} \quad (\text{S29})$$

using the phonon frequencies $\{\omega_{\alpha}\}$ calculated at the PBE level using the density functional perturbation theory²³ (DFPT) as implemented in QE. The adsorption energies, reaction energies, and barrier heights reported in Fig. M2A, M3D and M3F include the ZPE correction. The (inverse) reaction rates reported in Fig. M2B include the contribution from $A_{\text{vib}}(T)$.

To reduce the high computational cost of the DFPT calculations, we (i) use slab models that are smaller compared to those used in the electronic structure calculations and (ii) freeze a few layers from the bottom, as detailed in Sections S2.5 and S3.5 for the two surface systems. The D3 dispersion correction is not included in the DFPT calculations because QE does not support calculating its Hessian. However, we verified that all active phonon modes for non-transition state structures and all but one phonon modes for the transition state structures have positive frequencies.

S2 Water on the α -Al₂O₃(0001) surface

S2.1 Atomic structure

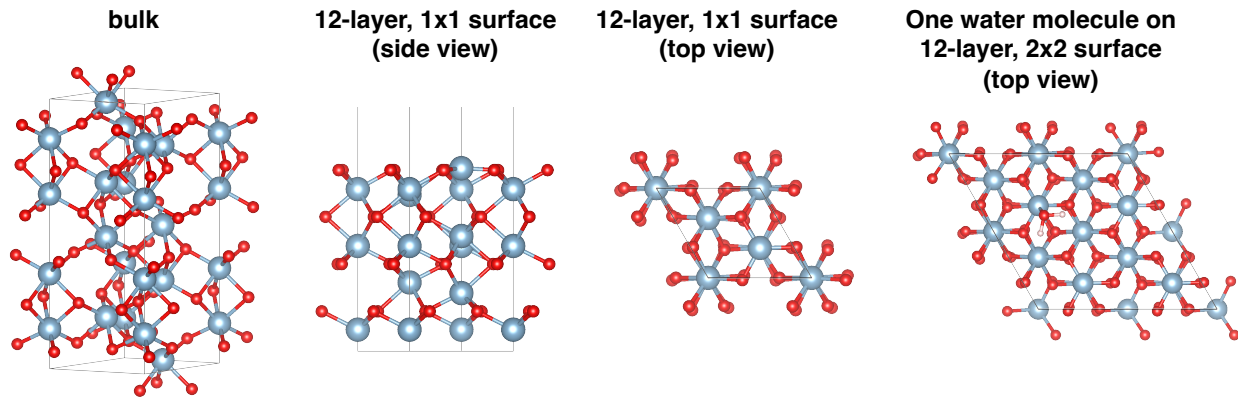


Figure S1: Atomic structure of bulk α -Al₂O₃, clean α -Al₂O₃(0001) surface, and water on 2×2 α -Al₂O₃(0001) surface.

The atomic structures of bulk α -Al₂O₃, clean α -Al₂O₃(0001) surface, and water on 2×2 α -Al₂O₃(0001) surface are shown in Fig. S1. Following conventions in literature,^{24–27} each atomic layer is counted as a layer in the slab model. This is different from the convention used for the rutile TiO₂(110) surface (see Section S3.1).

S2.2 Converged energetics

The final numbers for the ZPE-corrected adsorption energy E_{ads} , reaction energy ΔE , and barrier ΔE^\ddagger for the water dissociation on α -Al₂O₃(0001) surface reported in Fig. M2A are tabulated in Table S2. These numbers are obtained by combining the energetics from calculations with a finite vacuum size of 23 Å in the direction perpendicular to the slab and a correction to account for an infinite vacuum size (Δ_∞). As detailed in the rest of this section, these numbers have been converged to accuracy within 1 kcal/mol. The convergence with respect to basis size, slab/surface size, and Brillouin zone sampling, which gives rise to the numbers listed in column “23 Å vacuum” in Table S2, is detailed in Section S2.3. The correction to account for an infinite vacuum, which gives rise to the numbers in column “ Δ_∞ ”, is discussed in Section S2.4. The ZPE correction

evaluated in the harmonic approximation using PBE is discussed in Section S2.5.

S2.3 Slab size, surface size, basis size, and Brillouin zone sampling

We follow Ref. 24 and other previous DFT studies (e.g., Refs. 25–28) of the same system for the notation of slab size (in terms of the number of atomic layers) and surface size. We will use a notation (n, m) for a slab of n layers, with the top m layers being allowed to relax during geometry optimization. A vacuum of 23 Å perpendicular to the slab plane is used throughout this section. The determination of a correction to account for the infinite vacuum limit is detailed in Section S2.4.

PBE+D3 and SCAN+D3. By first fixing the slab size to be (12, 8), we determined from Table S3 the following converged parameters for the plane wave PBE+D3 calculations:

- a kinetic energy cutoff of 50 Hartree or 1350 eV,
- a surface size of 2×2 ,
- Brillouin zone sampling at the Γ point.

By fixing these parameters and scanning over different slab sizes, we determined from Table S4 a converged slab size of (12, 8). These parameters are also used for the SCAN+D3 calculations, except that we use a higher kinetic energy cutoff of 2000 eV due to the harder pseudopotential.

PBE0+D3 and MP2. By fixing the slab size to be (9, 5), we determined from Table S5 that a TZ Gaussian basis set is sufficient for converged energetics for both mean-field (HF and PBE) and correlated (MP2) calculations. The same table also confirms that the PBE+D3 energetics from PySCF with the TZ Gaussian basis set agrees with the plane wave PBE+D3 numbers from Quantum Espresso (the remaining difference, which is less than 1 kcal/mol, is mostly caused by the use of different pseudopotentials). By fixing the basis set and scanning over different slab sizes, we verified with data in Table S6 that the (12, 8) slab, which converges the PBE+D3 energetics as concluded from Table S4, is also sufficient to converge the MP2 energetics. Finally, we also expect

Table S3: PBE+D3 dissociation energy ΔE_{diss} and the barrier height ΔE^\ddagger (both in kcal/mol) of a single adsorbed water on the $\alpha\text{-Al}_2\text{O}_3(0001)$ surface evaluated for a (12, 8) slab model with different kinetic energy cutoff (in Hartree), surface size, and Brillouin zone sampling. The parameters that converge both energies within 1 kcal/mol are highlighted in grey.

K.E. cutoff	surface	k -point mesh	ΔE_{diss}	ΔE^\ddagger
50	2×2	$1 \times 1 \times 1$	-9.3	4.6
100	2×2	$1 \times 1 \times 1$	-9.3	4.6
100	2×2	$2 \times 2 \times 1$	-9.3	4.6
50	3×3	$1 \times 1 \times 1$	-9.0	4.6

Table S4: Same reaction energetics as in Table S3 for different choices of slab size. All other parameters are fixed at their optimum values determined in Table S3. The slab size that converge both energies within 1 kcal/mol is highlighted in grey.

Slab size	ΔE_{diss}	ΔE^\ddagger
(12,5)	-9.0	4.4
(12,7)	-9.2	4.6
(12,8)	-9.3	4.6
(15,10)	-9.4	4.6

these parameters (TZ basis set and (12, 8) slab) to be sufficient for PBE0+D3 calculations, as the PBE0 energy is a mixture of the PBE energy and the HF exchange energy.

Table S5: Reaction energetics (in kcal/mol) from HF, MP2, and PBE+D3 calculated using PySCF with the GTH-cc-pVXZ basis sets for a (9, 5)/ 2×2 surface model with Γ -point Brillouin zone sampling. A TZ basis set (highlighted in grey) is sufficient to converge reaction energetics to within 1 kcal/mol for both mean-field (HF, PBE, and hence also PBE0) and correlated (MP2) calculations. The basis-set converged PBE+D3 reaction energetics also agree with those from plane wave/Quantum Espresso very well (the small difference is likely caused by the difference in pseudopotentials).

Basis set	HF		MP2		PBE	
	ΔE_{diss}	ΔE^\ddagger	ΔE_{diss}	ΔE^\ddagger	ΔE_{diss}	ΔE^\ddagger
DZ/PySCF	-13.6	12.6	-8.4	7.2	-7.6	5.0
TZ/PySCF	-13.6	13.5	-8.7	7.2	-8.7	5.3
QZ/PySCF	-13.6	13.5	-8.7	7.2	-8.7	5.3
PW/QE					-9.0	4.7

LNO-CCSD(T). As detailed in Section S1.3, we use Eq. (S23) to evaluate the LNO-CCSD(T)

Table S6: Reaction energetics (in kcal/mol) from MP2 calculated for different slab models using a TZ basis set and for a 2×2 surface model with Γ -point Brillouin zone sampling. As in Table S4, the same (12, 8) slab model is sufficient to converge the MP2 energy.

Slab	ΔE_{diss}	ΔE^\ddagger
(12,5)	-9.1	7.1
(12,7)	-9.0	7.2
(12,8)	-9.1	7.0
(15,10)	-9.1	7.4

reaction energetics, which must be converged with respect to both the LNO truncation parameter, λ_o and λ_v , defined in Eq. (S8), and the “embedded” cluster size, M_o , defined in Eq. (S23). For the LNO truncation parameters, we follow Ref. 19 and scan λ_v while keeping $\lambda_o = 10\lambda_v$. For the embedded cluster size M_o , we note that the localized occupied orbitals (LOs) for this system fall into two classes: 4 LOs on the water molecule and 48 LOs for every layer of oxygen atoms (there are no LOs localized on aluminum due to the electron negativity difference between aluminum and oxygen). This results in a natural way of systematically increasing M_o :

1. LOs on the water molecule only, $M_o = 4$,
2. plus LOs on the first oxygen layer, $M_o = 4 + 48 = 52$,
3. plus LOs on the second oxygen layer, $M_o = 4 + 48 + 48 = 100$,
4. plus LOs on the third oxygen layer, $M_o = 4 + 48 + 48 + 48 = 148$,
5. plus LOs on the third oxygen layer, $M_o = 4 + 48 + 48 + 48 + 48 = 196 = N_o$.

The convergence with respect to M_o for different choices of λ_v is shown in Fig. S2 for a (12, 8)/ 2×2 surface model using a TZ basis set. For both the adsorption energy and the barrier height, convergence to an accuracy better than 1 kcal/mol is achieved using even the 4 LOs on the water molecule, while the reaction energy shows a slower convergence and requires further including the first oxygen layer (i.e., $M_o = 52$). Nevertheless, for the numbers reported in Table S2 and in the main text, we use the fully converged results from $M_o = 148$. The M_o -converged

reaction energies and barrier heights from different choices of λ_v (marked as stars in Fig. S2) are then plotted in Fig. M1E to investigate the convergence with respect to the LNO truncation, where we also see a fast convergence to the desired accuracy of 1 kcal/mol, as already discussed in the main text. The convergence with respect to surface size, slab thickness, and basis size is presented in Fig. M1G and M1H, from which we determine that the $(12, 8)/2 \times 2$ model and the TZ basis set are sufficient for converging both the reaction energy and the barrier height to chemical accuracy. These parameters are then used to evaluate the adsorption energy without further modifications.

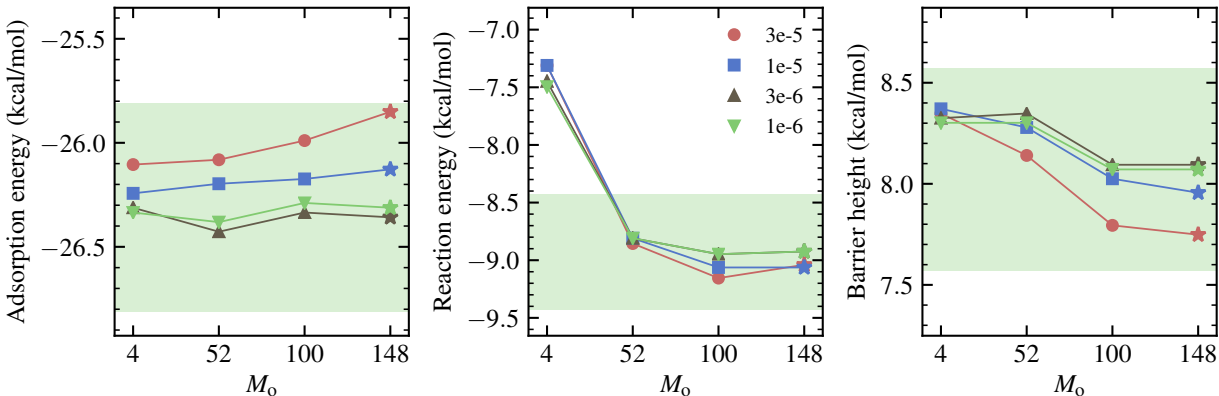


Figure S2: Convergence of the LNO-CCSD(T) adsorption energy (left), reaction energy (middle), and barrier height (right) for water- Al_2O_3 with respect to the embedded cluster size M_o . A $(12, 8)/2 \times 2$ surface model and a TZ basis set are employed. Data of different color correspond to different virtual LNO truncation parameter λ_v (the occupied truncation parameter is chosen to be $\lambda_o = 10\lambda_v$ in all cases). The range of ± 0.5 kcal/mol from the most converged number (i.e., $\lambda_v = 10^{-6}$ and $M_o = 148$) is highlighted by the green shaded area.)

S2.4 Infinite vacuum correction

PBE+D3, SCAN+D3, and PBE0+D3. We use the dipole correction at PBE+D3 level as implemented in Quantum Espresso to account for the finite-vacuum effect. For the $(12, 8)$ slab model, we found $\Delta_\infty = -0.62$, $+1.35$, and $+0.52$ kcal/mol for the adsorption energy, reaction energy, and barrier height, respectively, as also tabulated in Table S2. These corrections are also applied to obtain the final SCAN+D3 and PBE0+D3 energetics in Table S2.

MP2 and LNO-CCSD(T). For MP2, we manually increase the dimension perpendicular to the

surface (z_{\max}) and repeat the calculations using the (12, 8) model and the TZ basis set. The resulting energies are plotted in Fig. S3, where extrapolation to the infinite vacuum limit using the following function is also shown

$$E(z_{\max}) = \frac{A}{z_{\max} - B} + E(z_{\max} = \infty) \quad (\text{S30})$$

From the plot, we obtain $\Delta_{\infty} = -0.74, 2.05$, and 0.75 kcal/mol for E_{ads} , ΔE , and E^{\ddagger} , respectively, as tabulated in Table S2. In the spirit of the MP2 composite correction, these corrections are also applied to obtain the final LNO-CCSD(T) energetics in Table S2.

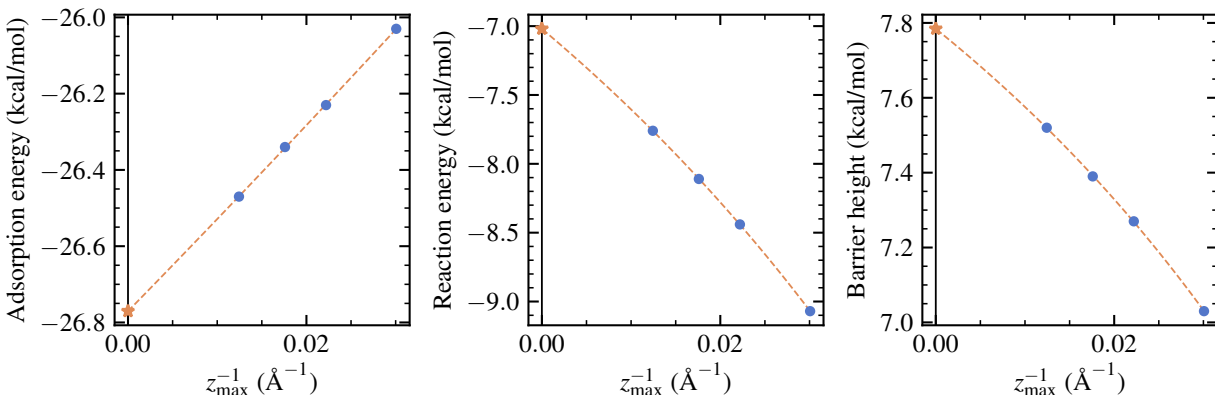


Figure S3: Convergence of the MP2 adsorption energy (left), reaction energy (middle), and barrier height (right) for water- Al_2O_3 as a function of the reciprocal dimension perpendicular to the surface. Extrapolations based on Eq. (S30) are shown in orange.

S2.5 Vibrational corrections

DFPT calculations at PBE level are performed for the clean surface, molecular adsorption, dissociated adsorption, and transition state of a $(9,5)/2 \times 2$ surface model. Atoms in the four layers from the bottom are kept frozen in the DFPT calculations to reduce the computational cost. The reciprocal space is sampled at the Γ -point. The ZPE corrections for the adsorption energy, reaction energy, and barrier height are already listed in Table S2. The temperature-dependent vibrational activation free energy used to evaluate the reaction rates shown in Fig. M2B is plotted in Fig. S4.

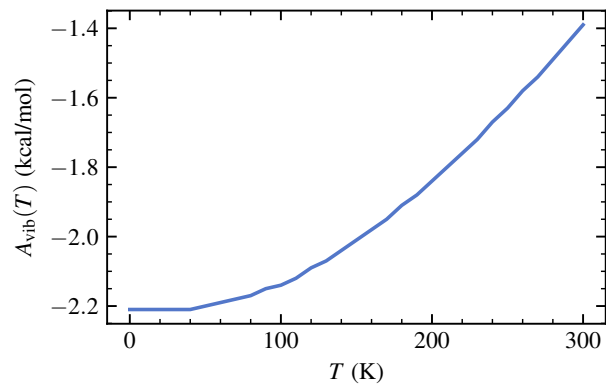


Figure S4: The temperature-dependent vibrational free energy for the barrier height of water dissociation on Al_2O_3 surface calculated according to Eq. (S29).

S3 Water on the rutile $\text{TiO}_2(110)$ surface

S3.1 Atomic structure

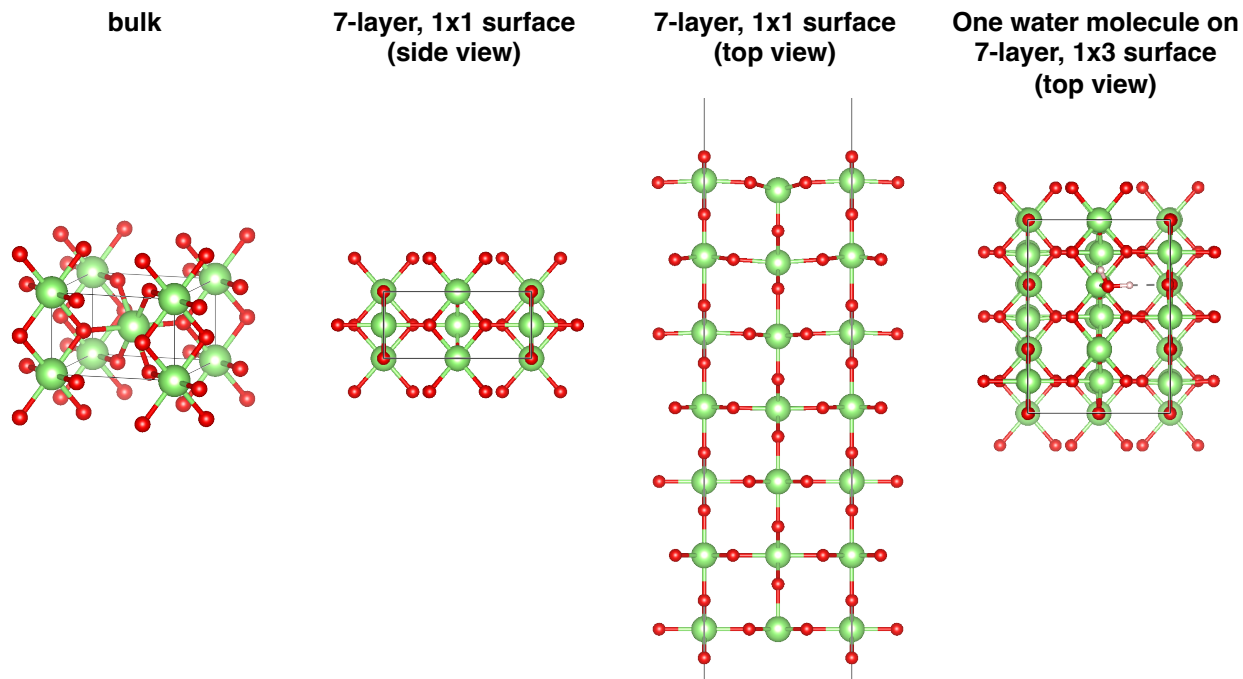


Figure S5: Atomic structure of bulk rutile TiO_2 , clean rutile $\text{TiO}_2(110)$ surface, and water on 1×3 rutile $\text{TiO}_2(110)$ surface.

The atomic structures of bulk rutile TiO_2 , clean rutile $\text{TiO}_2(110)$ surface, and water on 1×3 rutile $\text{TiO}_2(110)$ surface are shown in Fig. S5. The definition of atomic layer follows the literature,^{29–31} where an atomic layer includes a [Ti-O] plane plus the two O planes above and beneath it. Note that this is different from the convention used in Section S2.1 for water on $\alpha\text{-Al}_2\text{O}_3(0001)$ surface.

S3.2 Converged energetics

The final numbers for the ZPE-corrected adsorption energy E_{ads} , reaction energy ΔE , and barrier height ΔE^\ddagger for the water adsorption and dissociation on rutile $\text{TiO}_2(110)$ surface reported in Fig. M3D and M3F are tabulated in Table S7. These numbers are obtained by combining the ener-

getics from calculations with a finite vacuum size of 18 Å in the direction perpendicular to the slab and a correction to account for an infinite vacuum size (Δ_∞). As detailed in the rest of this section, these numbers have been converged to accuracy within 1 kcal/mol. The convergence with respect to basis size, slab/surface size, and Brillouin zone sampling, which gives rise to the numbers listed in column “18 Å vacuum” in Table S7, is detailed in Section S3.3. The correction to account for an infinite vacuum, which gives rise to the numbers in column “ Δ_∞ ”, is discussed in Section S3.4. The ZPE correction evaluated in the harmonic approximation using PBE is discussed in Section S3.5.

S3.3 Slab size, surface size, basis size, and Brillouin zone sampling

We follow Ref. 30 and other DFT studies on the same system for the notation of slab size and surface size. Like in water- Al_2O_3 , we use a notation (n, m) to denote a slab of n layers, with the top m layers being allowed to relax during geometry optimization

PBE+D3, PBE+ U +D3, and SCAN+D3. Our preliminary tests suggest that a kinetic energy cutoff of 1350 eV and a $2 \times 2 \times 1$ k -point mesh properly converge the basis set size and the Brillouin zone for the PBE+D3 reaction energetics. By fixing these parameters, we determined a converged surface size of 1×3 and a converged slab model of (7, 4) from Fig. S6. The same parameters are used for PBE+ U +D3 and SCAN+D3, except that we use a higher kinetic energy cutoff of 2000 eV for SCAN+D3 due to the harder pseudopotential.

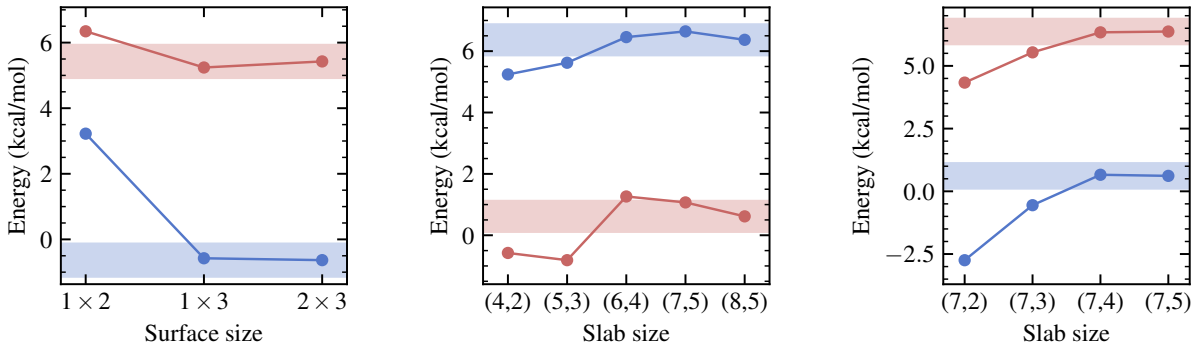


Figure S6: Convergence of the PBE+D3 reaction energy (blue) and barrier height (red) for water dissociation on TiO_2 surface with respect to the surface size (left), the total slab size (middle), and the active slab size (right). A (4, 2) slab model is used for the left panel, while a 1×3 surface model is used for the other two. All calculations were done with the Brillouin zone sampled by a $2 \times 2 \times 1$ mesh.

Table S7: Final numbers (highlighted in grey) for the ZPE-corrected adsorption energy E_{ads} , reaction energy ΔE , and barrier height ΔE^\ddagger for a single water molecule adsorption and dissociation on the rutile $\text{TiO}_2(110)$ surface calculated using different methods as reported in Fig. M3D and M3F. Columns labelled by “18 Å” list the regular total electronic energy contributions using a supercell with a vacuum of 18 Å in the z-direction (see Section S3.3). Columns labelled by “ Δ_∞ ” list the estimated infinite vacuum correction (see Section S3.4). Columns labelled by “ZPE” list the zero-point energy correction (see Section S3.5). All numbers are in kcal/mol.

Method	E_{ads}				ΔE				ΔE^\ddagger			
	18 Å	Δ_∞	ZPE	Final	18 Å	Δ_∞	ZPE	Final	18 Å	Δ_∞	ZPE	Final
PBE+D3	-22.63	0.20	1.40	-21.03	0.65	-0.03	-1.55	-0.93	6.26	-0.09	-2.82	3.35
PBE+U(1)+D3	-21.34	0.20	1.40	-19.74	0.42	-0.03	-1.55	-1.16	5.84	-0.09	-2.82	2.93
PBE+U(3)+D3	-18.60	0.20	1.40	-17.00	-0.16	-0.03	-1.55	-1.74	4.87	-0.09	-2.82	1.96
PBE+U(5)+D3	-15.64	0.20	1.40	-14.04	-0.91	-0.03	-1.55	-2.49	3.69	-0.09	-2.82	0.78
PBE+U(8)+D3	-10.88	0.20	1.40	-9.28	-2.42	-0.03	-1.55	-4.00	1.51	-0.09	-2.82	-1.40
PBE+U(11)+D3	-5.90	0.20	1.40	-4.30	-4.30	-0.03	-1.55	-5.88	-1.06	-0.09	-2.82	-3.97
SCAN+D3	-25.91	0.20	1.40	-24.31	1.31	-0.03	-1.55	-0.27	7.45	-0.09	-2.82	4.54
MP2	-25.65	0.51	1.40	-23.74	0.92	0.07	-1.55	-0.56	9.28	-0.15	-2.82	6.31
CCSD(T)	-22.74	0.51	1.40	-20.83	-0.41	0.07	-1.55	-1.89	7.06	-0.15	-2.82	4.09

MP2. Table S8 suggests that a TZ basis set is sufficient to converge reaction energetics evaluated at both mean-field (HF) and correlated (MP2) level. Table S9 suggests that the HF part of the MP2 energy requires sampling the Brillouin zone with a $2 \times 2 \times 1$ mesh, while the correlation part converges faster and requires only Γ -point Brillouin sampling. Ideally, we thus would like to use the following energy expression for the total MP2 energy

$$E_{\text{MP2}}^{(7,4)}(\text{TZ}/2 \times 2 \times 1) \approx E_{\text{HF}}^{(7,4)}(\text{TZ}/2 \times 2 \times 1) + E_{\text{MP2,c}}^{(7,4)}(\text{TZ}/1 \times 1 \times 1) \quad (\text{S31})$$

However, neither the TZ/2×2×1 HF calculation nor the TZ/1×1×1 MP2 calculation is feasible with the available computational resources. We thus approximate them using the following composite correction

$$\begin{aligned} E_{\text{HF}}^{(7,4)}(\text{TZ}/2 \times 2 \times 1) &\approx E_{\text{HF}}^{(7,4)}(\text{DZ}/1 \times 1 \times 1) \\ &+ \underbrace{E_{\text{HF}}^{(6,4)*}(\text{TZ}/1 \times 1 \times 1) - E_{\text{HF}}^{(6,4)*}(\text{DZ}/1 \times 1 \times 1)}_{\Delta E_{\text{HF}}^{(6,4)*}(\text{TZ}/1 \times 1 \times 1)} \\ &+ \underbrace{E_{\text{HF}}^{(6,4)*}(\text{DZ}/2 \times 2 \times 1) - E_{\text{HF}}^{(6,4)*}(\text{DZ}/1 \times 1 \times 1)}_{\Delta E_{\text{HF}}^{(6,4)*}(\text{DZ}/2 \times 2 \times 1)} \end{aligned} \quad (\text{S32})$$

$$\begin{aligned} E_{\text{MP2,c}}^{(7,4)}(\text{TZ}/1 \times 1 \times 1) &\approx E_{\text{MP2,c}}^{(7,4)}(\text{DZ}/1 \times 1 \times 1) \\ &+ \underbrace{E_{\text{MP2,c}}^{(6,4)*}(\text{TZ}/1 \times 1 \times 1) - E_{\text{MP2,c}}^{(6,4)*}(\text{DZ}/1 \times 1 \times 1)}_{\Delta E_{\text{MP2,c}}^{(6,4)*}(\text{TZ}/1 \times 1 \times 1)} \end{aligned} \quad (\text{S33})$$

where $(6,4)^*$ is derived from the $(7,4)$ slab model by removing one atomic layer from the bottom (this layer was kept frozen during the geometry relaxation). The final MP2 reaction energy and barrier height estimated using Eqs. (S31) to (S33) are shown in Table S10, along with all the energy components.

LNO-CCSD(T). We calculate the LNO-CCSD(T) correlation energy for the $(6,4)^*$ slab model using Eq. (S23) with a MP2 composite correction using the MP2 energy for the $(7,4)$ slab model calculated by Eq. (S31). The convergence of the LNO-CCSD(T) reaction energetics with respect

Table S8: Reaction energy ΔE and barrier ΔE^\ddagger of water dissociation on TiO_2 for a (4, 2) slab model with 1×2 surface calculated using MP2 with Γ -point Brillouin zone sampling and different Gaussian basis sets. Contributions from the HF and the MP2 correlation energy components to the MP2 energetics are also shown. As highlighted in gray, both the HF energy and the MP2 correlation energy contributions to the reaction energetics converge to accuracy within 1 kcal/mol with a TZ basis set.

Basis set	ΔE			ΔE^\ddagger		
	HF	MP2,c	MP2	HF	MP2,c	MP2
DZ	0.5	6.1	6.6	10.8	0.4	11.2
TZ	-1.1	7.0	5.9	9.9	-0.2	9.7
QZ	-1.4	7.3	5.9	9.7	0.4	10.1

Table S9: Same reaction energetics as in Table S8 but calculated using MP2 with a DZ basis set and different Brillouin zone sampling. Contributions from the HF and the MP2 correlation energy components to the MP2 energetics are also shown. As highlighted in gray, the HF energy converges to accuracy within 1 kcal/mol with $2 \times 2 \times 1$ Brillouin zone sampling, while the MP2 correlation energy converges even faster with Γ -point Brillouin zone sampling.

BZ sampling	ΔE			ΔE^\ddagger		
	HF	MP2,c	MP2	HF	MP2,c	MP2
$1 \times 1 \times 1$	0.5	6.1	6.6	10.8	0.4	11.2
$2 \times 2 \times 1$	-1.5	6.2	4.7	10.7	-0.1	10.6
$3 \times 3 \times 1$	-1.7			10.7		

Table S10: Estimating the MP2 reaction energy and barrier height for the (7,4) slab model using Eqs. (S31) to (S33).

	Slab	Basis set	BZ sampling	E_{ads}	ΔE	ΔE^\ddagger
HF	(7,4)	DZ	$1 \times 1 \times 1$	-15.8	-6.1	9.1
	(6,4)*	DZ	$1 \times 1 \times 1$	-15.6	-5.6	9.2
	(6,4)*	TZ	$1 \times 1 \times 1$	-18.6	-7.1	8.4
	(6,4)*	DZ	$2 \times 2 \times 1$	-7.7	-4.6	9.6
	(6,4)*	Δ TZ	$1 \times 1 \times 1$	1.2	-1.4	-0.8
	(6,4)*	DZ	$\Delta 2 \times 2 \times 1$	6.7	1.0	0.3
	(7,4)	TZ	$\Delta 2 \times 2 \times 1$	-10.9	-6.5	8.6
MP2,c	(7,4)	DZ	$1 \times 1 \times 1$	-17.6	6.0	1.9
	(6,4)*	DZ	$1 \times 1 \times 1$	-17.4	6.2	2.0
	(6,4)*	TZ	$1 \times 1 \times 1$	-17.7	7.7	0.8
	(6,4)*	Δ TZ	$1 \times 1 \times 1$	-0.4	1.4	-1.3
	(7,4)	TZ	$1 \times 1 \times 1$	-17.2	7.4	0.7
MP2	(7,4)	TZ	$2 \times 2 \times 1$	-25.6	0.9	9.3

to the cluster size M_o is shown in Fig. S7, where the choice of M_o follows that in water- Al_2O_3 , i.e., the first 4 LOs are localized on water, followed by 72 LOs in each atomic layer. Figure S7 suggests that for all choices of the LNO truncation parameters, $M_o = 292$, i.e., including LOs in up to the fourth atomic layer is sufficient for a converged results. The converged energetics from using different LNO truncation parameters (marked as stars in Fig. S7) are then plotted in Fig. M3C to investigate the convergence with respect to the LNO truncation, where we also see a fast convergence to the desired accuracy of 1 kcal/mol, as already discussed in the main text.

S3.4 Infinite vacuum correction

PBE+D3, PBE+ U +D3, and SCAN+D3. We use the dipole correction at PBE level as implemented in Quantum Espresso to account for the finite-vacuum effect. For the (7,4) slab model, we obtained $\Delta_\infty = 0.20$, -0.03 and -0.09 kcal/mol for the adsorption energy, reaction energy, and barrier height, respectively, as also tabulated in Table S7. These corrections are also applied to obtain the final PBE+ U +D3 and SCAN+D3 energetics in Table S7.

MP2 and LNO-CCSD(T). For MP2, we manually increase z_{max} and repeat the calculations using

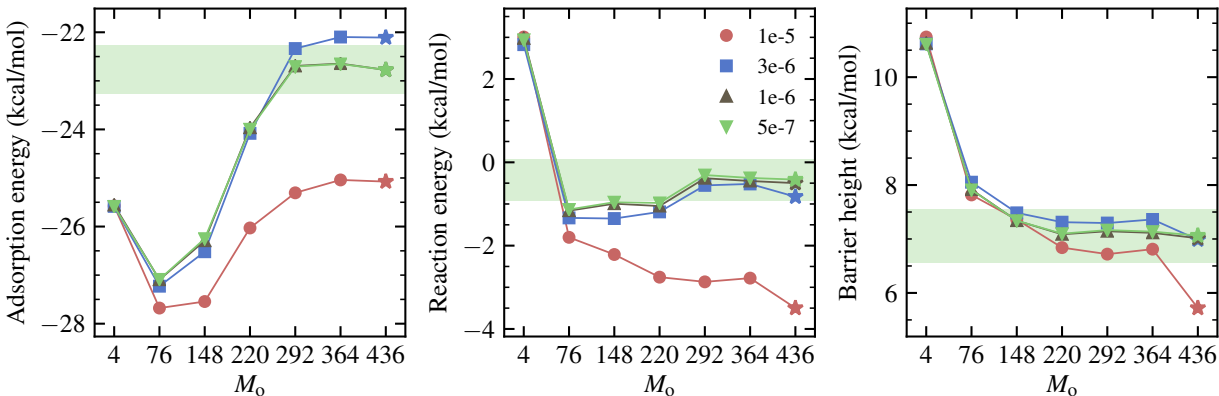


Figure S7: Convergence of the LNO-CCSD(T) adsorption energy (left), reaction energy (middle), and barrier height (right) for water-TiO₂ with respect to the embedded cluster size M_0 . The (6,4)^{*}/1 × 3 surface model and the TZ basis set are employed, with a MP2 composite correction evaluated using the (7,4)/1 × 3 surface model. Data of different color correspond to different virtual LNO truncation parameter λ_v (the occupied truncation parameter is chosen to be $\lambda_0 = 10\lambda_v$ in all cases except for $\lambda_v = 5 \times 10^{-7}$, where $\lambda_0 = 10^{-5}$). The range of ± 0.5 kcal/mol from the most converged number (i.e., $\lambda_v = 5 \times 10^{-7}$ and $M_0 = 436$) is highlighted by the green shaded area.)

the (6,4)^{*} model and the DZ basis set. The resulting energies are plotted in Fig. S8, where extrapolation to the infinite vacuum limit using Eq. (S30) is also shown. From the plot, we obtain $\Delta_\infty = 0.51, 0.07$, and -0.15 kcal/mol for E_{ads} , ΔE , and E^\ddagger , respectively, which are also listed in Table S7. In the spirit of the MP2 composite correction, these corrections are also applied to obtain the final LNO-CCSD(T) energetics in Table S7.

S3.5 Vibrational corrections

DFPT calculations at PBE level are performed for the clean surface, molecular adsorption, dissociated adsorption, and transition state of a (4,2)/1 × 3 surface model. Atoms in the three layers from the bottom are kept frozen in the DFPT calculations to reduce the computational cost. The reciprocal space is sampled at the Γ -point in the DFPT calculations. The ZPE corrections for the adsorption energy, reaction energy, and barrier height are already listed in Table S7.

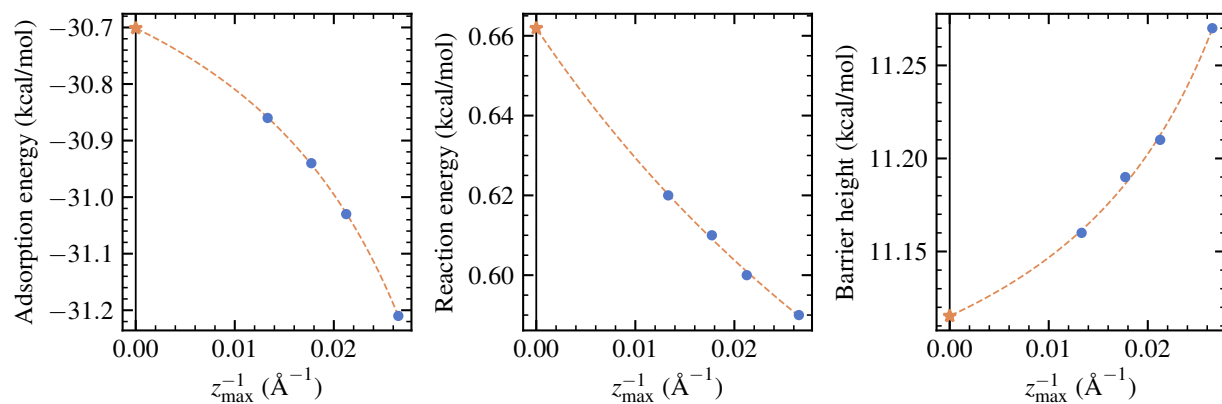


Figure S8: Convergence of the MP2 adsorption energy (left), reaction energy (middle), and barrier height (right) for water-TiO₂ as a function of the reciprocal dimension perpendicular to the surface. Extrapolations based on Eq. (S30) are shown in orange.

S4 Optimized geometries

S4.1 Bulk lattice constants

Table S11: Optimized lattice constants for bulk α -Al₂O₃ and rutile TiO₂. The PBE geometries are used for the subsequent surface calculations. PBE+D3 geometries are also included for comparison. For both crystals, the difference between PBE and PBE+D3 lattice constants is smaller than 0.4%. $a/b/c$ and $\alpha/\beta/\gamma$ are reported in Å and degree, respectively.

α -Al ₂ O ₃						
	a	b	c	α	β	γ
PBE	4.813	4.813	13.133	90	90	120
PBE+D3	4.796	4.796	13.085	90	90	120
Deviation	0.35%	0.35%	0.37%			
rutile TiO ₂						
	a	b	c	α	β	γ
PBE	4.649	4.649	2.971	90	90	90
PBE+D3	4.636	4.636	2.963	90	90	90
Deviation	0.28%	0.28%	0.27%			

S4.2 Optimized geometries

The PBE+D3 optimized surface geometries can be found in https://github.com/hongzhouye/supporting_data/tree/main/2023/arXiv%3A2309.14640.

S4.3 Energy uncertainty from the geometry

To estimate the uncertainty of the employed PBE+D3 geometry, we repeated the surface geometry relaxation using revPBE³²+D3 and compare the adsorption and reaction energetics evaluated using both the PBE+D3 optimized geometries and the revPBE+D3 optimized geometries. The results are listed in Table S12. We see that in both cases, the difference in the calculated adsorption and reaction energetics due to using the two sets of geometries is well-below chemical accuracy.

Table S12: revPBE+D3 adsorption and reaction energetics evaluated using PBE+D3-optimized geometries and revPBE+D3-optimized geometries. Results from the finite-vacuum geometries are shown without further ZPE or infinite vacuum corrections. All numbers are reported in kcal/mol.

	Geometry	E_{ads}	ΔE	ΔE^\ddagger
α -Al ₂ O ₃ (0001)	PBE+D3	-27.17	-10.23	5.39
	revPBE+D3	-27.22	-10.22	5.43
rutile TiO ₂ (110)	PBE+D3	-21.63	0.85	7.92
	revPBE+D3	-21.93	1.05	8.10

S5 Timing

The CPU cost for a typical single-point calculation using different methods for water-Al₂O₃ and water-TiO₂ is tabulated in Table S13 and Table S14.

Table S13: CPU cost (unit: hour) for a typical single-point calculation of the water-Al₂O₃ system measured using 16 CPU cores on a single node.

Method	Code	Basis	Slab model	k -point	CPU cost
PBE+D3	QE	100 Ry	(12, 8)/2 \times 2	1 \times 1 \times 1	0.3
SCAN+D3	QE	150 Ry	(12, 8)/2 \times 2	1 \times 1 \times 1	0.5
PBE0+D3	PySCF	TZ	(12, 8)/2 \times 2	1 \times 1 \times 1	42.2
HF	PySCF	TZ	(12, 8)/2 \times 2	1 \times 1 \times 1	30.6
MP2	PySCF	TZ	(12, 8)/2 \times 2	1 \times 1 \times 1	7.4
LNO-CCSD(T)	PySCF	TZ	(12, 8)/2 \times 2	1 \times 1 \times 1	403.4

Table S14: CPU cost (unit: hour) for a typical single-point calculation of the water-TiO₂ system measured using 16 CPU cores on a single node. For HF and MP2, CPU time for the most expensive calculation in the respective protocols is listed.

Method	Code	Basis	Slab model	k -point	CPU cost
PBE+D3	QE	100 Ry	(7, 4)/1 \times 3	2 \times 2 \times 1	4.7
PBE+ U +D3	QE	100 Ry	(7, 4)/1 \times 3	2 \times 2 \times 1	4.3
SCAN+D3	QE	150 Ry	(7, 4)/1 \times 3	2 \times 2 \times 1	49.0
HF	PySCF	DZ	(6, 4)*1 \times 3	2 \times 2 \times 1	1141.9
MP2	PySCF	TZ	(6, 4)*1 \times 3	1 \times 1 \times 1	113.7
LNO-CCSD(T)	PySCF	TZ	(6, 4)*1 \times 3	1 \times 1 \times 1	19099.9

References

- (1) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Corso, A. D.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. M. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **2009**, *21*, 395502.
- (2) Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Nardelli, M. B.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M.; Colonna, N.; Carnimeo, I.; Corso, A. D.; de Gironcoli, S.; Delugas, P.; DiStasio, R. A.; Ferretti, A.; Floris, A.; Fratesi, G.; Fugallo, G.; Gebauer, R.; Gerstmann, U.; Giustino, F.; Gorni, T.; Jia, J.; Kawamura, M.; Ko, H.-Y.; Kokalj, A.; Küçükbenli, E.; Lazzeri, M.; Marsili, M.; Marzari, N.; Mauri, F.; Nguyen, N. L.; Nguyen, H.-V.; de-la Roza, A. O.; Paulatto, L.; PoncÃ, S.; Rocca, D.; Sabatini, R.; Santra, B.; Schlipf, M.; Seitsonen, A. P.; Smogunov, A.; Timrov, I.; Thonhauser, T.; Umari, P.; Vast, N.; Wu, X.; Baroni, S. Advanced capabilities for materials modelling with Quantum ESPRESSO. *J. Phys. Condens. Matter* **2017**, *29*, 465901.
- (3) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*.
- (4) Sun, Q. Libcint: An efficient general integral library for Gaussian basis functions. *J. Comput. Chem.* **2015**, *36*, 1664–1671.
- (5) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: the Python-based simulations of chemistry framework. *Wiley Interdiscip. Rev. Comput. Mol. Sci* **2018**, *8*, e1340.

- (6) Sun, Q.; Zhang, X.; Banerjee, S.; Bao, P.; Barbry, M.; Blunt, N. S.; Bogdanov, N. A.; Booth, G. H.; Chen, J.; Cui, Z.-H.; Eriksen, J. J.; Gao, Y.; Guo, S.; Hermann, J.; Hermes, M. R.; Koh, K.; Koval, P.; Lehtola, S.; Li, Z.; Liu, J.; Mardirossian, N.; McClain, J. D.; Motta, M.; Mussard, B.; Pham, H. Q.; Pulkin, A.; Purwanto, W.; Robinson, P. J.; Ronca, E.; Sayfutyarova, E. R.; Scheurer, M.; Schurkus, H. F.; Smith, J. E. T.; Sun, C.; Sun, S.-N.; Upadhyay, S.; Wagner, L. K.; Wang, X.; White, A.; Whitfield, J. D.; Williamson, M. J.; Wouters, S.; Yang, J.; Yu, J. M.; Zhu, T.; Berkelbach, T. C.; Sharma, S.; Sokolov, A. Y.; Chan, G. K.-L. Recent developments in the PySCF program package. *J. Chem. Phys.* **2020**, *153*, 024109.
- (7) Cococcioni, M.; de Gironcoli, S. Linear response approach to the calculation of the effective interaction parameters in the LDA + U method. *Phys. Rev. B* **2005**, *71*, 035105.
- (8) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703–1710.
- (9) Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641–3662.
- (10) Hutter, J. New Optimization of GTH Pseudopotentials for PBE, SCAN, PBE0 Functionals. GTH Pseudopotentials for Hartree-Fock. NLCC Pseudopotentials for PBE. <https://github.com/juerghutter/GTH>, 2019.
- (11) Ye, H.-Z.; Berkelbach, T. C. Correlation-Consistent Gaussian Basis Sets for Solids Made Simple. *J. Chem. Theory Comput.* **2022**, *18*, 1595–1606.
- (12) Ye, H.-Z.; Berkelbach, T. C. Fast periodic Gaussian density fitting by range separation. *J. Chem. Phys.* **2021**, *154*, 131104.
- (13) Ye, H.-Z.; Berkelbach, T. C. Tight distance-dependent estimators for screening two-center and three-center short-range Coulomb integrals over Gaussian basis functions. *J. Chem. Phys.* **2021**, *155*, 124106.

- (14) Bintrim, S. J.; Berkelbach, T. C.; Ye, H.-Z. Integral-Direct Hartree-Fock and Møller-Plesset Perturbation Theory for Periodic Systems with Density Fitting: Application to the Benzene Crystal. *J. Chem. Theory Comput.* **2022**, *18*, 5374–5381.
- (15) Paier, J.; Marsman, M.; Hummer, K.; Kresse, G.; Gerber, I. C.; Ángyán, J. G. Screened hybrid density functionals applied to solids. *J. Chem. Phys.* **2006**, *124*, 154709.
- (16) Broqvist, P.; Alkauskas, A.; Pasquarello, A. Hybrid-functional calculations with plane-wave basis sets: Effect of singularity correction on total energies, energy eigenvalues, and defect energy levels. *Phys. Rev. B* **2009**, *80*, 085114.
- (17) Sundararaman, R.; Arias, T. A. Regularization of the Coulomb singularity in exact exchange by Wigner-Seitz truncated interactions: Towards chemical accuracy in nontrivial systems. *Phys. Rev. B* **2013**, *87*, 165122.
- (18) Rolik, Z.; Kállay, M. A general-order local coupled-cluster method based on the cluster-in-molecule approach. *J. Chem. Phys.* **2011**, *135*.
- (19) Rolik, Z.; Szegedy, L.; Ladjánszki, I.; Ladóczki, B.; Kállay, M. An efficient linear-scaling CCSD(T) method based on local natural orbitals. *J. Chem. Phys.* **2013**, *139*.
- (20) Nagy, P. R.; Kállay, M. Optimization of the linear-scaling local natural orbital CCSD(T) method: Redundancy-free triples correction using Laplace transform. *J. Chem. Phys.* **2017**, *146*.
- (21) Lehtola, S.; Jónsson, H. Pipek-Mezey Orbital Localization Using Various Partial Charge Estimates. *J. Chem. Theory Comput.* **2014**, *10*, 642–649.
- (22) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.
- (23) Baroni, S.; de Gironcoli, S.; Dal Corso, A.; Giannozzi, P. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.* **2001**, *73*, 515–562.

- (24) Hass, K. C.; Schneider, W. F.; Curioni, A.; Andreoni, W. The Chemistry of Water on Alumina Surfaces: Reaction Dynamics from First Principles. *Science* **1998**, 282, 265–268.
- (25) Hass, K. C.; Schneider, W. F.; Curioni, A.; Andreoni, W. First-Principles Molecular Dynamics Simulations of H₂O on α -Al₂O₃ (0001). *J. Phys. Chem. B* **2000**, 104, 5527–5540.
- (26) Ranea, V. A.; Carmichael, I.; Schneider, W. F. DFT Investigation of Intermediate Steps in the Hydrolysis of α -Al₂O₃(0001). *J. Phys. Chem. C* **2009**, 113, 2149–2158.
- (27) Wang, B.; Hou, H.; Luo, Y.; Li, Y.; Zhao, Y.; Li, X. Density Functional/All-Electron Basis Set Slab Model Calculations of the Adsorption/Dissociation Mechanisms of Water on α -Al₂O₃(0001) Surface. *J. Phys. Chem. C* **2011**, 115, 13399–13411.
- (28) Wirth, J.; Saalfrank, P. The Chemistry of Water on α -Alumina: Kinetics and Nuclear Quantum Effects from First Principles. *J. Phys. Chem. C* **2012**, 116, 26829–26840.
- (29) Harris, L. A.; Quong, A. A. Molecular Chemisorption as the Theoretically Preferred Pathway for Water Adsorption on Ideal Rutile TiO₂(110). *Phys. Rev. Lett.* **2004**, 93, 086105.
- (30) Lindan, P. J. D.; Zhang, C. Exothermic water dissociation on the rutile TiO₂(110) surface. *Phys. Rev. B* **2005**, 72, 075439.
- (31) Wang, Z.-T.; Wang, Y.-G.; Mu, R.; Yoon, Y.; Dahal, A.; Schenter, G. K.; Glezakou, V.-A.; Rousseau, R.; Lyubinetsky, I.; Dohnálek, Z. Probing equilibrium of molecular and deprotonated water on TiO₂(110). *Proc. Natl. Acad. Sci.* **2017**, 114, 1801–1805.
- (32) Zhang, Y.; Yang, W. Comment on “Generalized Gradient Approximation Made Simple”. *Phys. Rev. Lett.* **1998**, 80, 890–890.