Validating First Principles Molecular Dynamics Calculations of Oxide/Water Interfaces with X-ray Reflectivity Data (Supplemental Material)

Katherine J. Harmon,[1†] Kendra Letchworth-Weaver,[2,3†] Alex P. Gaiduk,[3] Federico Giberti,[3] Francois Gygi,[4] Maria K. Y. Chan,[2] Paul Fenter,[5] and Giulia Galli[3,6,7*]

[1]Applied Physics Graduate Program, Northwestern University, Evanston, IL 60208

[2]Center for Nanoscale Materials, Argonne National Laboratory, Lemont, IL 60439

[3]Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL 60637

[4]Department of Computer Science, University of California, Davis, CA 95616

[5]Chemical Sciences and Engineering Division, Argonne National Laboratory, Lemont, IL 60439

[6]Materials Science Division, Argonne National Laboratory, Lemont, IL 60439

[7]Department of Chemistry, University of Chicago, Chicago, IL 60637

[†]These authors contributed equally to this work.

[*]To whom correspondence should be addressed; E-mail: gagalli@uchicago.edu

I. First Principles Simulation Details

Each FPMD simulation consists of a fully-hydroxylated alumina slab with 6 oxygen layers along surface normal direction ($[O_3\text{-Al-Al}]_5\text{-}O_3$), 12 in-plane unit cells, and a water region with 96 water molecules, as shown for the PBE case in main text Figure 4. In total, these simulations included 120 Al atoms, 144 bulk O atoms, 72 surface O atoms (i.e., in the -Oh termination), 72 surface H atoms, 96 water O atoms, and 192 water H atoms (696 total atoms). The bulk Al_2O_3 structures were optimized by minimizing the stress tensor in DFT using a rhomobohedral 3 x $2\sqrt{3}$ x 1 supercell sampled at the gamma point (commensurate with a 3x3x1 k-point sampling of the conventional hexagonal unit cell with 12 Al atoms and 18 O atoms) with periodic boundary conditions and a cutoff energy of 60 Ry for PBE[1] calculations and 80 Ry for optB88[2] calculations. To form the interface, a rhombohedral supercell was constructed with a 3 x $2\sqrt{3}$ parallel reconstruction and 27.057 Å between periodic images, giving hexagonal bulk alumina lattice parameters of a = 4.81 Å and c = 13.16 Å for PBE, a = 4.80 Å and c = 13.13 Å for optB88 using a $3e^-$ pseudopotential ($3e^-$PP), and a = 4.77 Å and c = 13.04 Å using an $11e^-$PP (as described in Section II.C of the main text). The PPs were generated using PBE and were designed for use at an energy cutoff of 60 Ryd. FPMD runs were carried out a with cutoff energy 60 Ry at constant temperature (NVT) using a BDP thermostat [3] with T = 400 K for PBE simulations and T = 330 K for optB88 simulations. In each independent trajectory, the positions of the surface atoms and the water molecules were initialized from a randomly chosen snapshot from the classical MD simulations described in Section SII.A. For each simulation, four independent trajectories were sampled for various durations ranging from 2.5 – 10 ps each (see main text Table I) and were averaged to derive the final structure. FPMD electron densities were obtained by averaging 64 random snapshots for each independent trajectory for a total of 256 snapshots for each simulation.

II. Classical Molecular Dynamics Simulations

II.A Finite Size Effects

Due to the computational cost of FPMD, we turn to classical MD to study effects of the simulation size (number of atoms present) and water height. Our CMD simulations were performed using the LAMMPS [4] code and the ClayFF [5] and SPC/E[6] forcefields to represent the Al_2O_3 interface and water, respectively. Calculations were carried out for alumina slabs with 6-24 oxygen layers along the Al_2O_3(0001) surface normal direction and 96-2520 water molecules. Calculations with 6 alumina layers and 96 water molecules were carried out in the NVT ensemble from which trajectories were extracted for our FPMD simulations. In CMD calculations with ≥ 280 water molecules, the cell was first optimized via NPT and then trajectories were carried out with NVT.

These simulations rely on empirical force fields and were performed as a preliminary evaluation of size effects that can guide choices for the more expensive FP simulations. As CMD predicts atomic and not electron distributions, we compared computed and measured X-ray reflectivity (XR) data only through the use of predicted atomic number densities and tabulated form factors (FFs) following Eq. 2b in the main text. The results of the comparisons with the measured XR intensities are summarized in Table SI.

From the results of the NPT-NVT calculations (6-24 layers Al_2O_3 layers and 280 - 2520 water molecules), we determined that the use of a 6-layer alumina slab for the FPMD calculations is sufficiently large so that the central portion of the slab (which is used to model the "bulk" substrate structure factor in the XR intensity comparisons) is unaffected by the system size, i.e., in terms of surface relaxations. The predicted Al_2O_3 slab lattice parameter for the CMD 6-layer slab in the NPT-NVT ensemble is 2.2227 Å (+2.6% relative to the experimental value of 2.1663 Å),

TABLE SI. Summary of χ^2 values from classical MD calculations and comparison with the experimental result. The number (#) of layers in the Al_2O_3 slab, number of water molecules (N_w), and number of parallel unit cells in the Al_2O_3 slab (N_{UC}) are indicated. The simulation temperature T is given in Kelvins. ΔV is the volume change of the simulation cell via compression along alumina (001)-direction relative to the volume used in the PBE simulation (see main text). Simulations were run either in the NVT ensemble or first with NPT and then NVT (NPT-NVT). d_{006} is the Al_2O_3 slab lattice parameter and $d_{Al\text{-}Al}$ is the inter-layer distance between adjacent Al atoms at the center of the alumina slab. χ^2 is calculated using both atomic and ionic form factors (FF).

# Layers	# Waters (N_w)	# || UC (N_{UC})	N_w/N_{UC}	T (K)	ΔV (%)	Ensemble	d_{006} (Å)	$d_{Al\text{-}Al}$ (Å)	χ^2	
									Atomic FF	Ionic FF
6	96	12	8	300	0	NVT	2.2551	0.4858	101	55
6	96	12	8	300	10	NVT	2.2516	0.4866	90	55
6	280	12	23.3	300	0	NPT-NVT	2.2228	0.5422	38	80
6	588	24	24.5	300	0	NPT-NVT	2.2227	0.5421	37	77
12	580	12	48.3	300	0	NPT-NVT	2.2345	0.5200	22	55
18	880	12	73.3	300	0	NPT-NVT	2.2366	0.5175	28	66
24	2520	24	105	300	0	NPT-NVT	2.2375	0.5152	25	55
Expt.							2.1663	0.5230	1.54	227[a]

[a]Calculated from the best fit atomic structure (Figure 2c) but with ionic FF instead of atomic FF in the model XR calculation (Eq. 2c).

which is in closest agreement with the experimental value among the CMD calculations (Table SI). A larger number of layers in the Al_2O_3 slab is associated with a decreased accuracy of the lattice constant prediction, which manifests as an increasing displacement of the Al_2O_3 atomic layer heights relative to the experimental best fit model (Figure S1a). However, the XR intensity calculated from the 6-layer NPT-NVT simulations also results in the worst χ^2 level of agreement with the XR data (χ^2 = 38 using atomic FF; Table SI) despite having the most accurate lattice parameter. It is important to note that because the XR signal is computed using reciprocal lattice units *L* (r.l.u.), there is a cancellation of errors related to the larger slab lattice parameter from the center of the CMD slab. Therefore, the substantial inaccuracies in the bulk-like region depicted in Figure S1a do not directly affect χ^2 as much as might be expected, but they do indirectly affect χ^2

due to an effective scaling of the bulk-like internal aluminum coordinate at the center of the slab, $d_{Al\text{-}Al}$, and a scaling of the water structure. This cancellation of errors may partially explain why the χ^2 is better for the simulations with 12-24 layers compared with 6 layers, despite their less accurate prediction of the slab lattice parameter d_{006}. The χ^2 accuracy appears to correlate, in part, with $d_{Al\text{-}Al}$. There may also be contributions from the water structure to the tabulated χ^2 values, as discussed below. Overall, a 6-layer slab structure is more accurate with regard to some metrics but not others, and it is more computationally tractable in FPMD than a 12-layer or larger slab.

We observe that the interfacial water structure does not change significantly by including more than 280 water molecules. The electron density distributions of the interfacial water for all NPT-NVT simulations (Figure S1b) are similar with only subtle changes in the peak density of the first hydration layer. The peak density monotonically decreases as the number of water molecules per surface unit cell increases, but the change amounts to only ~5%. Unfortunately, it is difficult to quantify the effects of an increasing number of water molecules on the computed XR signal

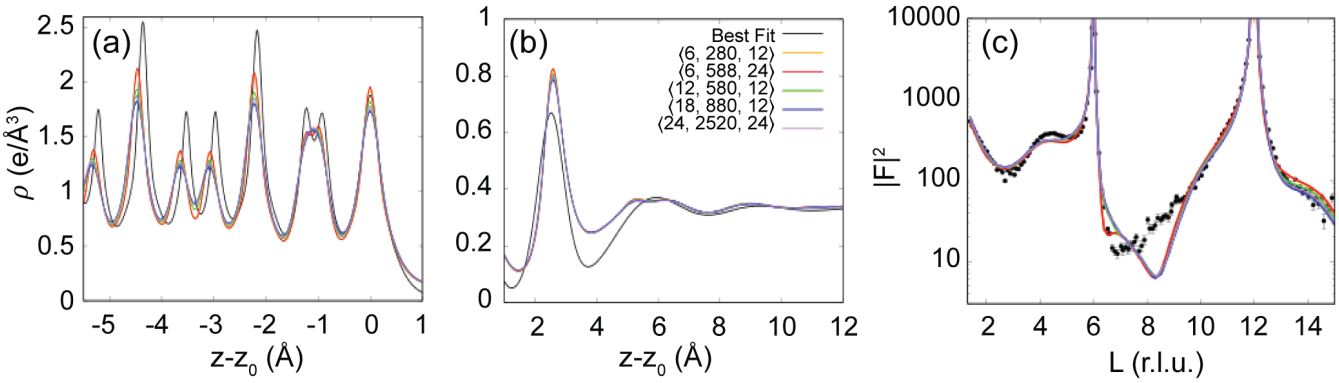

FIG S1. Size-dependent structural variations using classical MD (CMD) simulations. The electron density distributions $\rho(z)$ as a function of perpendicular distance to the surface layer of oxygen (at $z - z_0 = 0$) shows (a) the Al_2O_3 interface deviates more from the experimental best fit (black line) as the number of slab layers increases while (b) the water distribution is largely conserved as the number of water molecules is increased. (c) The computed XR signals $|F|^2$ as a function of reciprocal lattice unit L (r.l.u.) give agreements of $22 < \chi^2 < 38$ with the experimental data. CMD curves are designated by $\langle \#Al_2O_3$ layers, $\#H_2O$, #|| unit cells$\rangle$ (see Table SI). The electron density distributions are calculated from the predicted elemental density distributions and convolved with atomic FF (main text Eq. 3a).

(Figure S1c) and resulting χ^2 agreement with the XR data because of overwhelming errors in the bulk Al_2O_3 prediction (as we found in the main text). Our FPMD simulations with 96 water molecules are larger than [7, 8] or in line with [9] the number of water molecules included in other recent FPMD calculations of the Al_2O_3/water system, but size effects cannot be eliminated completely, as we find in the main text.

II.B Water Height Effects on XR Intensities

The FPMD XR intensities show a much lower intensity at L = 6-9 than is present either in the XR data or in the CMD XR intensities (main text Figure 5). Notably, the CMD interfacial water height Δ_w was 2.55 Å whereas the FPMD interfacial water heights were 2.74 Å for PBE and 2.68 Å for optB88, and the experimental result was 2.52 Å. This suggested that the water height might be a contributing factor in the appearance of the low-intensity region of the computed FPMD XR signals.

To test the effect of the interfacial water height Δ_w, we calculated the XR signals after we rigidly shifted the CMD water structure away from the alumina surface. We carried out this test using the simulated density profiles from the 6-layer Al_2O_3 with 12 parallel unit cells and 96 H_2O molecules to be consistent with the model used in our FP calculations. We note that in addition to the interfacial water height, we consider the ratio of the water height to the lattice spacing, Δ_w/d_{006}, to account for the differences in lattice constant predictions between FP (Δ_w/d_{006} = 1.25 for PBE) and CMD (Δ_w/d_{006} = 1.13). Indeed, as the water height is increased, a similar low intensity region to that observed in the FPMD calculations is observed from CMD (Figure S2). However, the magnitude of the dip is not as large as that in the FP XR intensities. When the CMD water height is displaced to the PBE interfacial water height of ~2.7 Å, only a minor reduction in intensity is

observed. When the CMD water height is displaced to 2.82 Å above the surface such that Δ_w/d_{006}=1.25 (as predicted by PBE) the intensity still does not drop as low as it is in the PBE calculation, and its shape is different but qualitatively much closer to that of the PBE XR intensity. Thus, while the water structure, including its height above the alumina, appears to play a part in the low scattering intensity at L = 6-9, the water height alone does not explain all of the observed discrepancy. Lastly, the interplay between predicted water structure and predicted bulk structure for a given simulation is an important consideration.

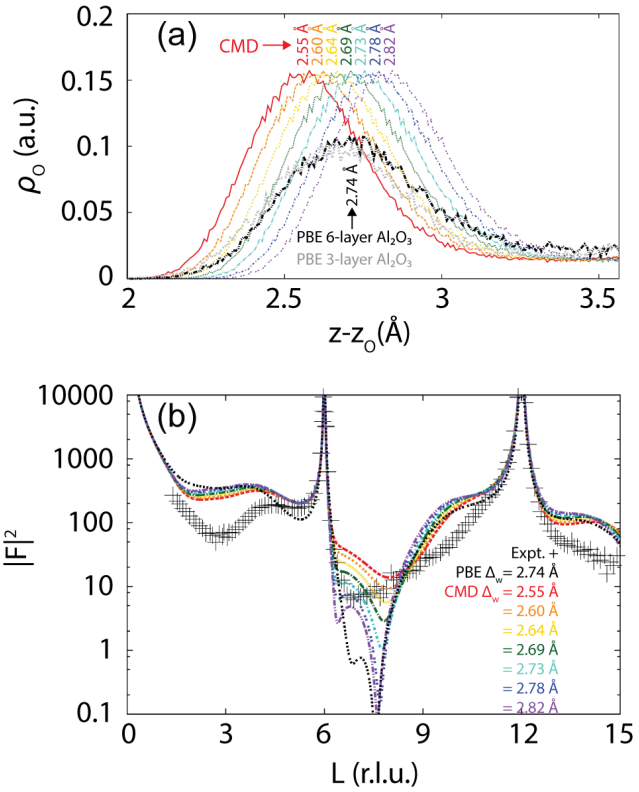

FIG S2. Effects of the displacement of the CMD water height. (a) The oxygen atomic densities ρ_O from PBE and CMD at several water heights are given as the distance between the alumina surface oxygen and the peak water oxygen of the first hydration layer (z-zO). (b) Larger water heights Δ_w in CMD yield a lower XR intensity $|F|^2$ at the midzone ($6 < L < 9$) but do not precisely reproduce the PBE result.

III. Sensitivity to Valence Electron Distributions

III.A Effect of Pseudopotential

Here we explore the source of the apparent electron delocalization observed around the Al cores in the Al_2O_3 slabs for optB88 simulations with an $11e^-$PP but not with the $3e^-$PP. We performed several calculations from a single snapshot using PBE (with both the $3e^-$PP and $11e^-$PP) and the hybrid function PBE0 [10] (with the $11e^-$PP) and compared to the optB88 simulations with $3e^-$PP and $11e^-$PP to test how the delocalization varies with these exchange correlation functional/PP configurations. Figure S3a,b show that for a given PP, the delocalization error is nearly identical for these functionals. In particular, the broadening of the electron density around the Al peaks in the Al_2O_3 is present to the same extent for all $11e^-$PP calculations, showing that the use of hybrid and non-local exchange correlation functionals to treat the valence electrons does not remedy the broadening of the charge distribution. Thus, the error results from the use of the PP itself.

Next we investigate the Al core FF used to calculate the total electron densities shown in Figures S3a,b. The core FF for the $3e^-$PP and $11e^-$PP (Figure S3c) vary significantly at small momentum transfers, q, where the zero-q value is equal to the magnitude of charge on the ionic core (i.e., 10 core electrons and 3 valence for the $3e^-$PP versus 2 core electrons and 11 valence for the $11e^-$PP). As a reference, with atomic FF, the zero-q value is equal to the atomic number Z = 13. Conversely, the high-q structure of the $3e^-$ and $11e^-$ FF are identical, indicating that the 1s core electrons are treated in the same way for both PP. Thus, the differences in the total electron density distributions do not appear to relate to the 1s frozen core.

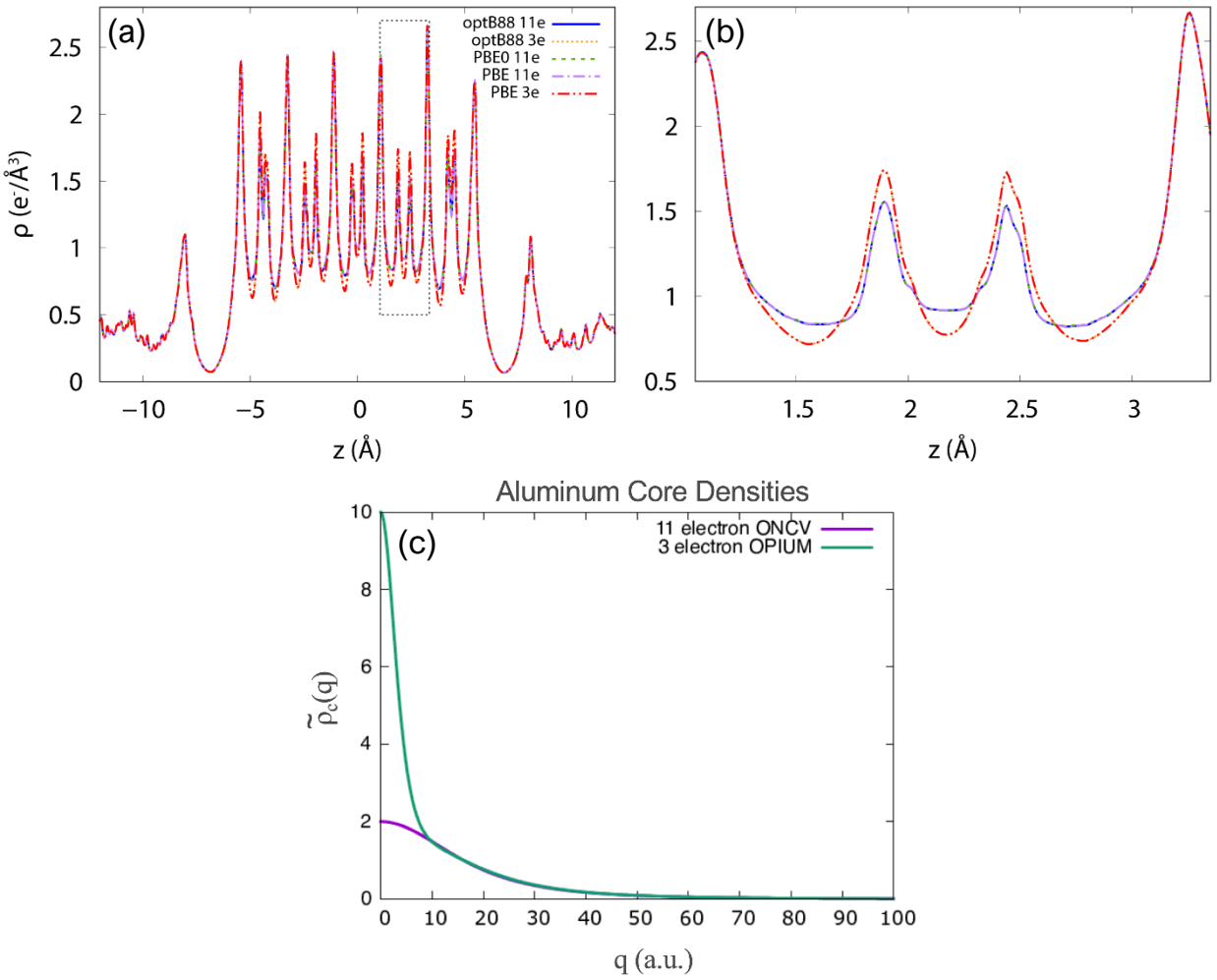

FIG S3. Effects of pseudopotential (PP) choice on electron delocalization in the Al_2O_3 slab, calculated using PBE, optB88, and PBE0. (a) Total electron densities $\rho(z)$, including the Al_2O_3 slab centered at z = 0 Å and the water regions, from a single MD snapshot show similar structures with equivalent atomic layer positions for all calculations (optB88 11e$^-$PP in blue and 3e$^-$PP in gold; PBE0 11e$^-$PP in green; PBE 11e$^-$PP in purple and 3e$^-$PP in red). (b) A close-up view of the region highlighted by the dotted box in (a) shows that the exchange correlation functional used to describe the valence electrons (PBE, optB88, or PBE0) does not affect the electron delocalization and that this phenomenon results from the use of the 11e$^-$PP. (c) The Al core form factors for a 3e$^-$ PP and an 11e$^-$PP vary significantly for $q < 10$ (i.e., the semicore electron states).

III.B Effect of Ionic Form Factors

In the majority of our calculations using FF, we assume neutral atomic FF. However, the DFT calculations show a low electron density at the bonding region between the Al_2O_3 surface and adjacent water, which may be indicative of charge transfer. Therefore, we performed additional calculations of the electron densities and XR intensities using the predicted elemental distributions from all simulations discussed in the main text but now using ionic FF (Al^{3+}, O^{2-}, and H^{+}). The results of the ionic FF calculations are compared to those using atomic FF in Table SII. The use of ionic FF to describe the electron distributions yields an improvement in χ^2 from PBE (χ^2 = 100) to optB88 with the 3e$^-$PP (χ^2 = 77) to optB88 with the 11e$^-$PP (χ^2 = 64). Further improvement is observed for the optB88 11e$^-$PP simulation with 10% compression of the unit cell (χ^2 = 45).

The electron density distributions for the 11e optB88 with 10% compression using the simulated elemental number density profile and four different valence electron configurations (those from atomic FF, ionic FF, the 11e$^-$PP, and the 3e$^-$PP) are compared in Figure S4. They are generally similar, but with subtle variations in both the Al_2O_3 slab (Figure S4b) and bonding region between the Al_2O_3 and water regions (Figure S4c), as we saw in the main text Figure 7. The electron density using ionic FF is in closer qualitative agreement with the experimental density in the solid-water bonding region than when using atomic FF (Figure S4c).

We observe that the relative accuracy of calculations using ionic versus atomic FF appears to correlate with the Al-Al spacing in the bulk Al_2O_3 (Figure S4e). Ionic FF consistently provide better agreement with the XR data for simulations in which the Al atoms are separated by $\leq$ 0.5131 Å (the PBE value), and atomic FF provide better agreement when this distance is > 0.5131 Å (Table SII and Figure S4e). When the Al atoms are closer together as in the 11e$^-$PP optB88

TABLE SII. χ^2 from first principles MD calculations and experiment using atomic and ionic form factors (FF). PBE and optB88 indicate the energy functionals used in the simulations. The pseudopotential PP includes either 3 or 11 valence electrons chosen for Al. Also noted are the volume change, ΔV, of the simulation cell via compression along alumina (001)-direction relative to the volume used in the PBE simulation; the total simulation time, *t*, after averaging over 4 independent trajectories; the Al_2O_3 lattice parameter, d_{006}; and the internal Al-Al coordinate at the center of the alumina slab, $d_{Al\text{-}Al}$.

Exchange Correlation	**PP**	**ΔV (%)**	**t (ps)[a]**	**d_{006} (Å)**	**$d_{Al\text{-}Al}$ (Å)**	**χ^2**	
						Atomic FF	**Ionic FF**
PBE	3	0	40	2.2028	0.5131	102	100
optB88	3	0	12.8	2.1982	0.5190	66	77
	11	0	11.6	2.1843	0.4986	100	64
	11	10	13.2	2.1809	0.4999	69	45
Experiment				2.1663	0.5230	1.54	227[a]

[a]Calculated from the best fit atomic structure (Figure 2c) but with ionic FF instead of atomic FF in the model XR calculation (Eq. 2c).

calculation with 10% compression ($d_{Al\text{-}Al}$ = 0.4999 Å), using atomic FF decreases the χ^2 accuracy because it leads to an excess electron density between the Al atoms (though not to the same extent we see when using the 11e^-PP electron density prediction directly). This excess electron density is counteracted by using ionic FF (Figure S4b). Conversely, when the Al-Al distance is sufficiently large as in the 3e^-PP simulation ($d_{Al\text{-}Al}$ = 0.5190 Å) and the experiment ($d_{Al\text{-}Al}$ = 0.5230 Å), the use of ionic FF decreases the accuracy because it underestimates the inter-layer electron density; this error can be rectified by using atomic FF. These comparisons reinforce the sensitivity of XR data to subtle variations in valence electron distributions. They also highlight the importance of choosing a suitable combination of exchange correlation functional and PP that will accurately predict both the interfacial water structure and the substrate solid structure (including its atomic structure and valence electron configuration). Recall that the experimental XR analysis makes use of atomic FF but does not include hydrogen in the model because of the weak scattering of X-rays

from protons, which is why even with atomic FF (which place an e$^-$ at the position of H atoms), the XR best fit density is small in the solid-water bonding region.

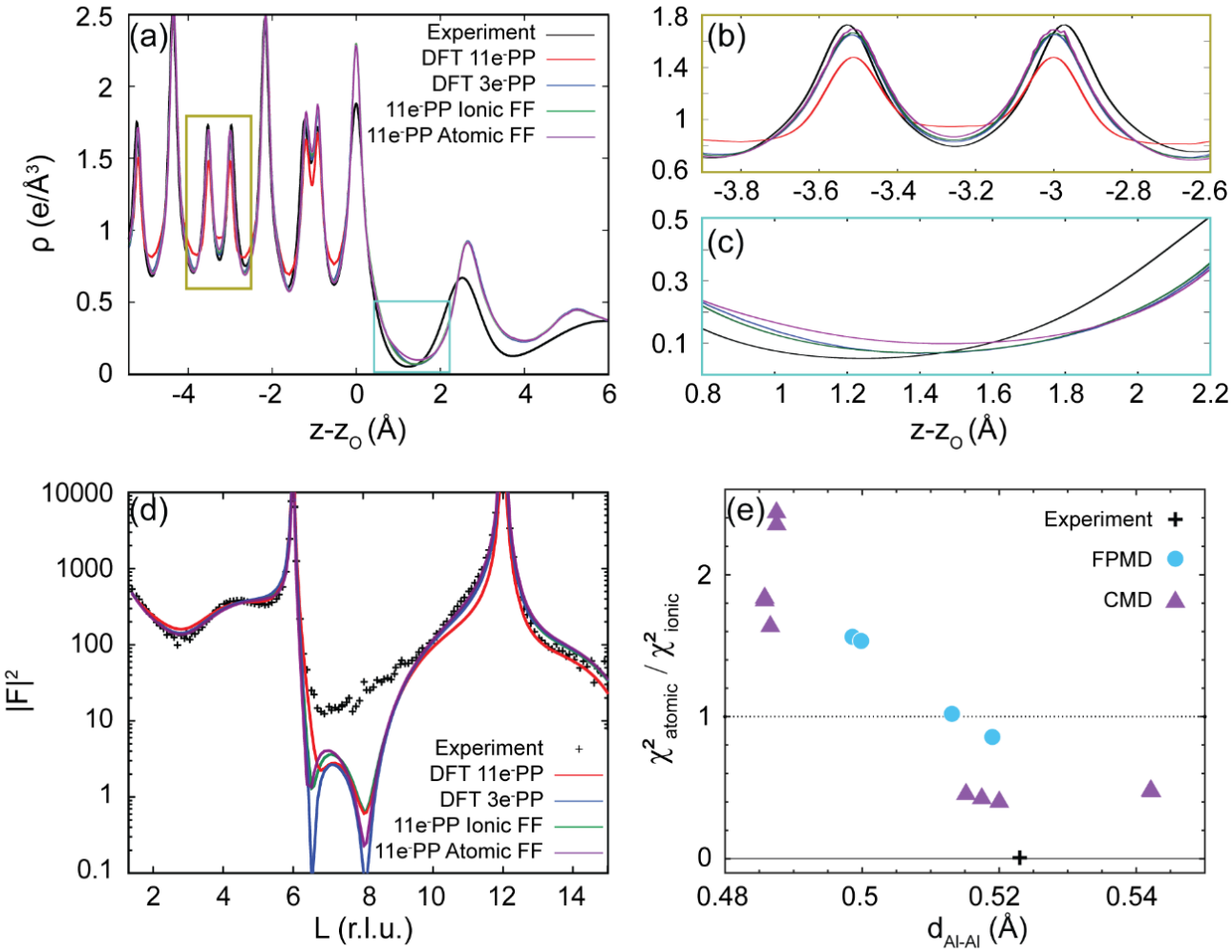

FIG S4. Effect on the XR signal of different electron density configurations with a comparison to ionic FF demonstrated using the atomic density prediction from the optB88 simulation with 10% compression and 11e$^-$PP. (a) Total electron density $\rho(z)$ referenced to the position of the alumina surface oxygen at z_O. (b) Close-up of bulk-like Al-Al electron density in the Al_2O_3 slab (gold box in (a)) showing a clear broadening of the charge distribution for the 11e$^-$PP. (c) Close-up of the electron-density of the solid-water bonding region (cyan box in (a)) showing close agreement in the electron densities between the 3e$^-$ and 11e$^-$PP and the ionic FF calculations. (d) Comparison of computed XR signals $|F|^2$ versus reciprocal lattice unit L (r.l.u.). (e) Trends in the relative accuracy of XR signals computed using atomic or ionic FF for first principles MD (blue circles) and classical MD (purple triangles) indicate that atomic FF outperform ionic FF ($\chi^2_{atomic}/\chi^2_{ionic} < 1$) when the predicted Al-Al distance $d_{Al\text{-}Al}$ exceeds 0.5131 Å (see Table SI for CMD simulation data and Table SII for FPMD results); the experimental result ('+') is shown for reference. The "DFT 11e$^-$PP," "DFT 3e$^-$PP," and "11e$^-$PP atomic FF" curves are the same data as those presented in the main text Figure 7.

IV. Additional Statistical Sampling

IV.A 11e$^-$PP optB88-10% with Atomic FF

The present analysis considers the same four independent trajectories obtained for the compressed optB88 simulations investigated in Section III.D of the main text but using atomic FF instead of the 3e$^-$PP to describe the electron distribution around the atomic centers. The atomic FF electron density (Figure S5a) is similar to that using the 3e$^-$PP (main text Figure 8), and the differences between each trajectory appear mostly in the interfacial water region (Figure S5b), as we saw with the 3e$^-$PP analysis. However, when using atomic FF, there are greater statistical deviations seen within the Al_2O_3 slab than we saw when using the 3e$^-$PP, with deviations within the slab in excess of one standard deviation from the mean appearing mainly at the position of Al atoms. With atomic FF, the mean variance of the four trajectories from their average ($\langle s_i(z) - \langle s(z) \rangle \rangle$ is $5.60\times10^{-11} \pm 0.0163$ e$^-$/Å^3, whereas with the 3e$^-$PP the mean variance is $1.91\times10^{-8} \pm 0.0165$ e$^-$/ Å^3. The levels of agreement with the XR data using atomic FF are slightly worse than those using the 3e$^-$PP and range from $57 < \chi^2 < 91$ ($\chi^2_{avg} = 69$). The simulated XR intensities deriving from the atomic FF electron densities (Figure S5c) show minor differences compared to the XR intensities calculated using the 3e$^-$PP, namely in the low-intensity region, $6 < L < 8$. The low-L XR signals with atomic FF (Figure S5d) are similar to those using the 3e$^-$PP.

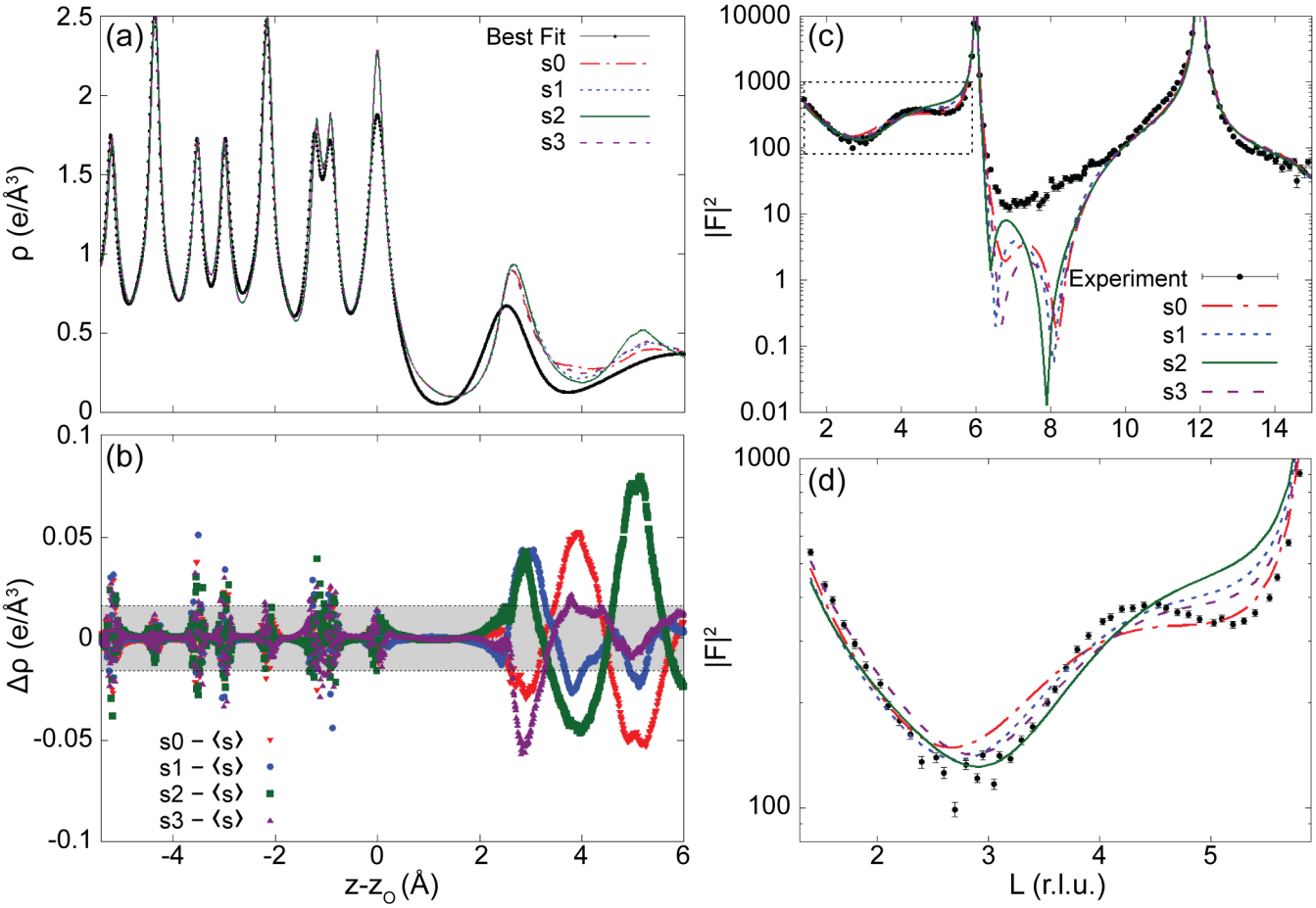

FIG S5. Role of sampling from four independent trajectories, s0-s3 (s0 red: s1: blue, s2: green, s3: purple), using the atomic density distribution from the optB88 simulation with 10% compression and atomic FF. (a) Total electron density distributions $\rho(z)$ referenced to the position of the alumina surface oxygen at z_O. (b) The difference in density, $\Delta\rho$, between each trajectory and the average of all four trajectories, $\langle s\rangle$ (which gives $\chi^2 = 69$); density differences for s0 - $\langle s\rangle$ are shown as red inverted triangles, for s1 - $\langle s\rangle$ as blue circles, for s2 - $\langle s\rangle$ as green squares, and for s3 - $\langle s\rangle$ as purple triangles. The total variance for all trajectories averages $5.60\times10^{-11} \pm 0.0163$ e$^-$/Å^3 (gray shaded box). (c) Simulated XR intensities $|F|^2$ versus reciprocal lattice unit L (r.l.u.) from each trajectory and (d) close-up view of the XR signals below the first Bragg peak (dotted box in (c)).

IV.B 3e$^-$PP PBE

For each trajectory considered individually the quality of fit ranges over $107 < \chi^2 < 133$ while the XR signal computed from the average electron densities from all four trajectories yields χ^2=119 (main text Table I). While the simulated intensities of the four trajectories are nearly indistinguishable at most momentum transfers, they vary by up to an order of magnitude for L =

7-8 (Figure S6a). Trajectories s1 and s3 (Figure S6b) exhibit the largest visual differences in the computed XR intensities but have similar elemental number density distributions. The average difference between s1 and s3 is found to be $\Delta n = 1/N \sum|n_1(z) - n_2(z)|$ = 0.0058 $\pm$ 0.0065 atoms/Å^3 while the maximum deviation occurs at the position of the Al_2O_3 surface oxygen ($\Delta n(z - z_0 = 0)$ = 0.0481 atoms/Å^3, or ~6% of the peak density of that layer). The interfacial water region shows the largest cumulative differences with an average deviation of Δn ($z \geq 0$) = 0.0076 $\pm$ 0.0069 atoms/Å^3.

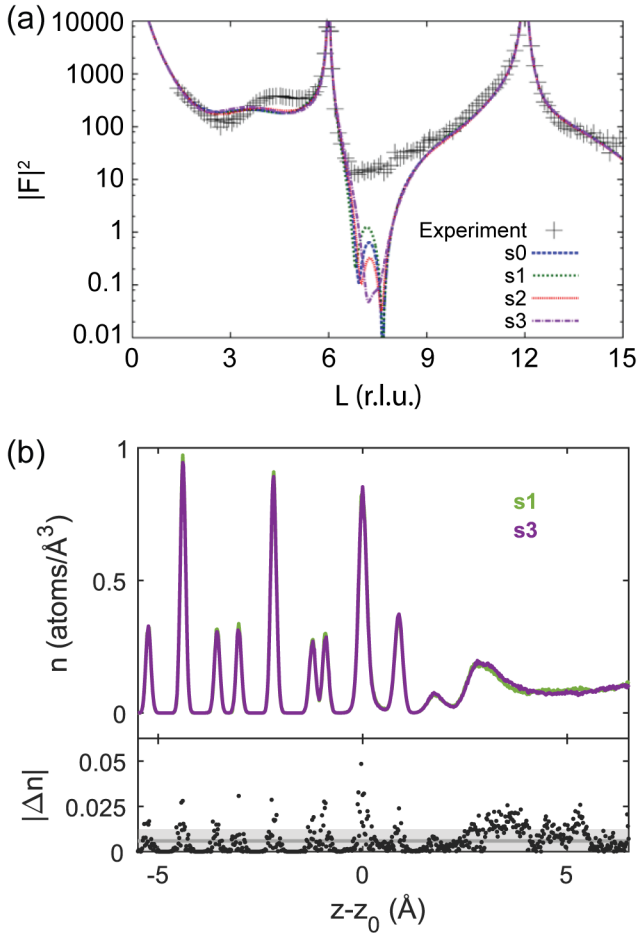

FIG S6. Role of sampling from four independent PBE trajectories, s0-s3 (s0: blue, s1: green, s2: red, s3: purple). (a) Comparison of computed XR intensities $|F|^2$ versus reciprocal lattice unit L (r.l.u.) shows a range of agreement with the experimental data with 107< χ^2 <133. (b) Top: atomic density profiles n(z) from trajectories s1 and s3 are nearly indistinguishable despite showing substantial differences in computed XR intensities; Bottom: the magnitude of the difference in atomic densities |Δn| (dots) averages 0.0058 $\pm$ 0.0065 atoms (gray line and shaded area).

V. Empirical Fitting of Simulation Structures

IV.A Parameterization of Simulated Structures

The simulation structures are translated into the least-squares fitting procedure that is used to determine a best fit structure to the experimental XR data, as described in detail in Ref. [8]. To do this, we must fit the simulated number density profiles of the solid alumina interface to Gaussians with parameters defining the atomic layer positions, rms widths, and coverages following the model described by Eqs. 2c and 3b in the main text. These parameters can then be optimized in order to minimize the errors associated with the solid interface and obtain the best structure (i.e., in terms of the χ^2 quality of agreement with the experimental data) that is possible with the given simulated water electron density profile.

The fitted PBE-predicted structure is given in Figure S7a and shows that the atomic layers of the alumina solid are well-described by Gaussians. The same process is carried out for the optB88 and compressed optB88 simulations, not shown. To ensure that we have correctly parameterized the alumina interface, we embed the simulated structure between a bulk alumina substrate and a bulk water structure in two ways – 1) the total effective electron density from the simulation, including the alumina surface and the interfacial water region, is directly placed between the bulk regions without any parameterization; and 2) the alumina interface is parameterized as described above, and the effective electron density of the water region only (which has not been parameterized) is directly embedded above it (Figure S7b-d). The two are compared in terms of the consistency between the density profiles and in terms of the χ^2 level of agreement of the computed XR intensities with the XR data (Figure S7e-g). The results show high consistency between the original (i.e., embedded effective densities directly from FPMD) and parameterized structures with some minor differences in the computed XR as a result of small

errors in the parameterization of the r.m.s. widths. This is due to the fact that the predicted number density profiles, though well-described by Gaussians, are not perfectly represented as such.

The bulk Al_2O_3 in the empirical analysis of the simulated structures and shown in Figure S7 is defined using the known experimental lattice constant of c = 12.998 Å (d_{006} = 2.1663 Å), an Al-Al distance of $d_{Al\text{-}Al}$ = 0.523 Å, and using atomic FF. Calculations were also carried out with the known bulk structure and ionic FF, which resulted in worse agreement with the experimental data, or with the slab lattice structures (i.e., as extracted from the middle layer of the simulated alumina slabs) and atomic FF, which gave a superior level of agreement for PBE but worse agreement for the optB88 calculations. The χ^2 values are summarized in Table SIII.

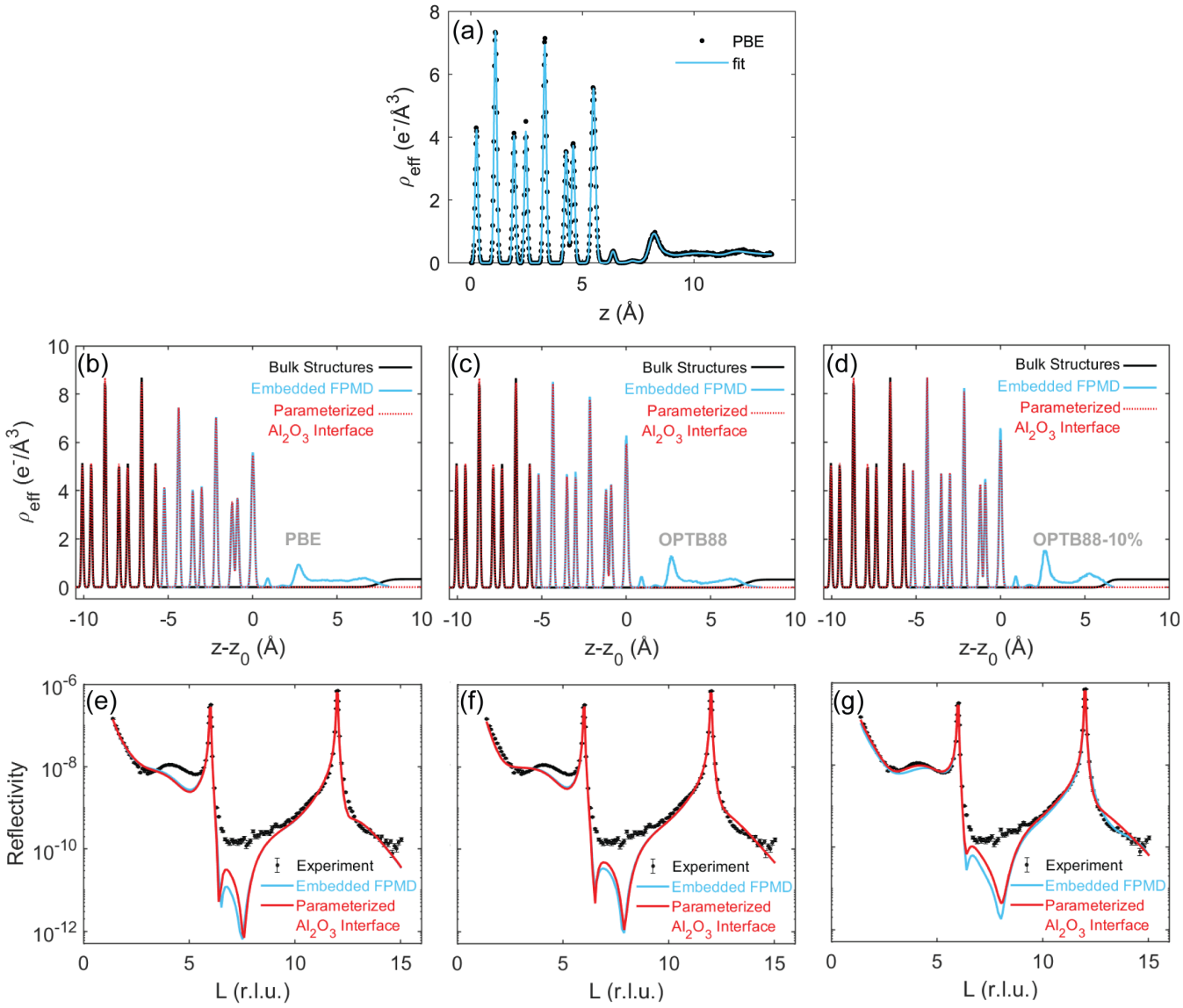

FIG S7. Embedded first-principles MD simulations for empirical fitting. (a) The PBE density profile is fit to a series of Gaussians to extract layer positions and widths in order to translate the FPMD prediction to the parameter space of the least squares fitting routine. Equivalent fits were carried out for optB88 and optB88-10% (not shown). The (b) PBE, (c) optB88, and (d) optB88-10% interfacial effective electron densities ρ_{eff} (see main text Eq. 3b) are calculated in two ways: 1) the simulated interface, including alumina and water regions, are directly embedded between a bulk substrate (with experimental bulk lattice constant of 2.1663 Å and Al-Al inter-layer distance 0.523 Å) and bulk water structure; or 2) the alumina surface is parameterized via the procedure shown in (a) and placed above the bulk alumina while the FPMD water structure is embedded between the parameterized surface and bulk water. The XR intensities versus reciprocal lattice unit L (r.l.u.) calculated from the fully embedded and parameterized (e) PBE, (f) optB88, and (g) optB88-10% density profiles show minor differences as a result of the parameterization of the Al_2O_3 surface. Densities and XR intensities calculated from the embedded structures are shown in blue; calculations with the parameterized alumina interface are in red. XR intensity calculations use atomic FF.

Table SIII. Initial χ^2 values of embedded structures from first principles molecular dynamics (FPMD) either with the full interface from FPMD (i.e., Al_2O_3 surface and interfacial water) embedded between the alumina bulk and bulk water (called "Embedded") or with the Al_2O_3 surface parameterized and the FPMD water embedded between alumina and bulk water (called "Parameterized"). χ^2 was calculated using atomic or ionic form factors (FF); with the embedded structures and atomic FF, χ^2 was calculated using an alumina bulk lattice constant d_{006} equal to the FPMD slab lattice parameter or the known experimental value of 2.1663 Å.

Exchange Correlation	PP	ΔV (%)	d_{006} (Å)	Atomic FF		Ionic FF
				Embedded	Parameterized	Embedded
PBE	3	0	2.2028	104	-	-
			2.1663	112	124	126
optB88	11	0	2.1843	105	-	-
			2.1663	83	90	88
		10	2.1809	75	-	-
			2.1663	54	51	73

To create a continuous electron density distribution at the boundaries of the embedded FPMD structures, error function windows are applied on the boundaries of the substrate, the FPMD, and the bulk water. The total electron density is given by

$$\rho(z) = \rho_{subst}(z) \times \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{z - z_s}{\sigma_s}\right)\right] + \rho_{\mathrm{FPMD}}(z) \times \frac{1}{4}\left[1 + \mathrm{erf}\left(\frac{z - z_s}{\sigma_s}\right)\right] \times \left[1 - \mathrm{erf}\left(\frac{z - z_w}{\sigma_w}\right)\right] + \rho_w \times \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{z - z_w}{\sigma_w}\right)\right]. \quad (1)$$

Here, $\rho_{subst}(z)$ is the electron density of the parameterized solid alumina substrate with the complementary error function, 1 – erf, applied to the "top" end of the substrate at $z = z_s$ where z_s is the edge of the substrate, and σ_s is the r.m.s. edge width of erfc; $\rho_{\mathrm{FPMD}}(z)$ is the electron density of the embedded FPMD region with an error function applied to the "bottom" end at the boundary between the solid substrate and FPMD water region ($z = z_s$, width = σ_s) and the complementary

error function applied to the "top" end at the boundary between the FPMD water and bulk water at $z = z_w$ and with r.m.s. edge width σ_w; ρ_w = 0.33 $e^-/Å^3$ is the density of bulk water described as an error function profile extending infinitely in the positive-z direction from z_w and with rms edge width σ_w.

The position of the error function step edges and their rms widths were found to have a non-negligible effect on χ^2. Therefore, these parameters were also optimized as a function of the parameters z_i and σ_i ($i = s, w$). Table II in the main text includes the χ^2 values for the structures after optimizing both the Al_2O_3 surface and the error function parameters. We note that there were several parameter values that led to lower χ^2 but which were excluded because they did not provide a robust evaluation the FPMD accuracy:

(1) For PBE and optB88 (without reduction of the supercell z-dimension), χ^2 could be reduced by truncating the embedded FPMD water structure at some z within the first hydration layer (i.e., by increasing z_s or σ_s such that a portion of the first peak was removed) or by truncating the FPMD structure in the region of the second hydration layer (by decreasing z_w or increasing σ_w such that a portion of the second peak was removed). Such structures then only partially describe the FPMD predictions.

(2) We determined that in the optB88 calculations, the χ^2 could be improved by significantly reducing z_s to the point of partially including the electron density of the alumina surface oxygen. This then leads to an overlap with the density in the surface layer from the parameterized interface and effectively double counts a portion of the electron density there.

(3) Also in the case of the optB88 calculation, certain combinations of z_w and σ_w that produced a sharp step edge between the FPMD water region and the bulk water region also

led to an improvement in χ^2. The abrupt step results from an erf cutoff that extends beyond the end of the simulation. Such cases were determined to produce non-physical electron density profiles.

Altogether, a series of rules for the error function positions and edge widths were devised to enable a physical and accurate assessment of the predicted water densities: 1) the included section of the FPMD simulations must be sufficiently separate from the parameterized alumina interface so as to avoid double counting of electron densities; 2) the onset of the FPMD water region must include the complete first hydration layer, i.e., including the full width of the electron density distribution of this layer; 3) the second hydration layer peak of the simulated FPMD cannot be truncated by the error function window; and 4) the boundary between the FPMD water and bulk water must present a physical, continuous transition from one to the other.

In principle, some of these concerns should also be important for the transition between the interfacial and bulk water regions for the direct FPMD calculations computed via Eq. 5 in the main text. In those calculations, we used the experimental bulk water electron density ρ_w = 0.33 $e^-/Å^3$ for the bulk water structure factor calculations, which creates a subtle discontinuity at the joining between the FPMD simulation cell and bulk water, as discussed above and in Section II.C. However, the choice to use the experimental bulk was made because the simulation cells were at times too small to contain the second hydration shell completely, and using the density at the center of the liquid region as a semi-infinite extension led to spurious predictions of χ^2 as well as substantial deviations between individual trajectories at the same simulation conditions. This means that the bulk water region is described accurately in all calculations of XR signal, but the asymptotic behavior of the water approaching bulk in the interfacial region may not be. To remedy this concern in future, more accurate simulations should:

(1) Select a sufficiently large unit cell FPMD unit cell such that the water is bulk-like at the center of the liquid region and, therefore, a semi-infinite extension of the simulation electron density can be used, as was done for the Al_2O_3 bulk. This was not feasible for the FPMD calculations in this work given the computational limitations.

(2) Conduct separate FPMD simulations of bulk water to determine bulk water density for a given theory at a given pressure (e.g. optB88 at 330 K under atmospheric pressure) and use this value. Unfortunately, obtaining the pressure of the water in the FPMD simulations of the interface is quite technically challenging, similar to conducting NPT simulations, and was not feasible in the present study.

To address the discontinuity that arises from the use of the experimental bulk with the simulated interfacial water density, we tested error function smoothing between these two regions, as was done for the empirical fitting procedure. However, in contrast to in the results of the empirical fitting analysis, we found that the impact of error function smoothing on χ^2 was negligible compared to using a sharp transition between the interfacial region and bulk water. This difference in χ^2 sensitivity likely arises because errors from the other regions in the calculation (e.g. bulk alumina) overwhelmed this feature when we did not control for such errors carefully through fitting.

IV.B Understanding χ^2 Results of Optimized Structures

Here we evaluate additional features of the interfacial density distributions after optimizing the alumina surface in order to understand the resulting χ^2 accuracy (see main text Table II and Figure 9). The water structure was calculated in two ways: 1) as the electron density distribution direct from the first principles simulations and 2) as the number density distribution dressed with

atomic FF. In particular, we aim to understand why the XR intensity computed using the water electron density distribution directly from PBE yields higher accuracy (χ^2 = 5.6) than that calculated by using the PBE-derived element density distribution dressed with atomic FF (χ^2 = 11.2) and also why the accuracy is similar to that of the compressed optB88 (optB88-10%) calculations (χ^2 = 5.6 and 5.5 using the simulated water electron density and the simulated number density, respectively). The optB88 calculations (before volume reduction) give even worse χ^2 accuracies after minimizing substrate errors: χ^2 = 14.8 when directly calculated using the simulated water electron density distribution and χ^2 = 17 when calculated with the water number densities and FF. Here we explore the details of the second hydration layer and of the Al_2O_3 surface relaxations to determine if they are associated with the observed χ^2 values, but we ultimately determined that these χ^2 differences correlate with the r.m.s. width of the Al_2O_3 surface oxygen layer, as detailed in the main text.

The side-by-side comparison of the water densities calculated directly from the simulations or from number densities and FF (main Figure 9a) reveals that there are negligible differences in the distributions of the second hydration layers between the two methods of calculating the electron density. Moreover, the PBE and optB88 second hydration layers are qualitatively more similar to one another and to the experimental best fit than those in the optB88-10% calculation. In addition, the distance between the first and second hydration layers are also similar for PBE and optB88, but the distance is shorter in the optB88-10% case. Finally, the optB88-10% structure shows a low-density region between the first and second hydration layers ($z \sim 3.5$) that is most similar to the low-density region found experimentally. However, the PBE water electron density distribution does not predict a similarly low density between the first and second water layers and yet gives an equivalent χ^2. These observations do not follow the trends seen in χ^2, and we

conclude that the details of the second hydration layer structures cannot explain the observed χ^2 values.

Next, we evaluate the alumina surface relaxation, which arises as the surface atoms adjust in order to compensate for errors in the water structure. The alumina surface oxygen layer relaxes outward from the bulk in all cases with the greatest shift in PBE and the smallest shift in optB88-10%. Moreover, PBE shows the greatest distortion of the top unit cell (the surface-most Al-Al-O_3) from the best fit structure with all three of these layers relaxing away from the bulk in PBE. In contrast, for both optB88 with and without compression and for the experimental best fit model, the Al atom next to the surface O relaxes inward. One might expect that because the optB88 and optB88-10% simulations are in closer qualitative agreement with the experimental best-fit structure, they would result in a more accurate χ^2 than PBE, but this is not the case.

We also found that the alumina surface relaxation was independent of how the water structure was calculated (i.e., using number densities and atomic FF or electron distributions directly from DFT). This further suggests that the surface relaxations cannot explain the optimized χ^2 values, in particular between the PBE calculations using the simulated water electron density or the simulated number density. We quantify the relaxation patterns in terms of the r.m.s. deviation from the relaxation profile found in the experimental best fit model (with χ^2=1.54). For PBE, the optimized Al_2O_3 relaxation patterns deviated from the experiment by 0.070 $\pm$ 0.003 Å (number density) and 0.072 $\pm$ 0.003 Å (electron density); for uncompressed optB88, these deviations were 0.042 $\pm$ 0.003 Å (number density) and 0.043 $\pm$ 0.003 Å (electron density); and for optB88-10%, these deviations were 0.025 $\pm$ 0.004 Å (number density) and 0.027 $\pm$ 0.003 Å (electron density). We conclude that the Al_2O_3 surface relaxation behaviors cannot explain why the PBE calculation using all DFT electrons for the water structure gives a better χ^2 than the one

using the water number density and atomic FF or why it is equivalent to the optimized optB88-10% results.

V. References

1. J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple*, Phys. Rev. Lett. **77**, 3865 (1996).
2. K. Jiří, R. B. David, and M. Angelos, *Chemical accuracy for the van der Waals density functional*, J. Phys. Condens. Matter **22**, 022201 (2010).
3. G. Bussi, D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*, J. Chem. Phys. **126**, 014101 (2007).
4. S. Plimpton, *Fast Parallel Algorithms for Short-Range Molecular Dynamics*, J. Comput. Phys. **117**, 1 (1995).
5. R. T. Cygan, J.-J. Liang, and A. G. Kalinichev, *Molecular Models of Hydroxide, Oxyhydroxide, and Clay Phases and the Development of a General Force Field*, J. Phys. Chem. B **108**, 1255 (2004).
6. H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *The missing term in effective pair potentials*, J. Phys. Chem. **91**, 6269 (1987).
7. M. DelloStritto, and J. Sofo, *Bond Polarizability Model for Sum Frequency Generation at the Al_2O_3(0001)–H_2O Interface*, J. Phys. Chem. A **121**, 3045 (2017).
8. K. J. Harmon, Y. Chen, E. J. Bylaska, J. G. Catalano, M. J. Bedzyk, J. H. Weare, and P. Fenter, *Insights on the Alumina–Water Interface Structure by Direct Comparison of Density Functional Simulations with X-ray Reflectivity*, J. Phys. Chem. C **122**, 26934 (2018).
9. P. Huang, T. A. Pham, G. Galli, and E. Schwegler, *Alumina(0001)/Water Interface: Structural Properties and Infrared Spectra from First-Principles Molecular Dynamics Simulations*, J. Phys. Chem. C **118**, 8944 (2014).
10. C. Adamo, and V. Barone, *Toward reliable density functional methods without adjustable parameters: The PBE0 model*, J. Chem. Phys. **110**, 6158 (1999).