Development of Classical Molecule–Surface Interaction Potentials Based on Density Functional Theory Calculations: Investigation of Force Field Representability.

Karen Johnston,*,† Claudia R. Herbers,‡ and Nico F. A. van der Vegt‡

Max-Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany, and Center of Smart Interfaces – Technische Universität Darmstadt, Petersenstr. 32, 64287 Darmstadt, Germany

E-mail: johnston@mpip-mainz.mpg.de

This is a peer-reviewed, accepted author manuscript of the following article: Johnston, K., Herbers, C. R., & van der Vegt, N. F. A. (2012). Development of classical molecule-surface interaction potentials based on density functional theory calculations: investigation of force field representability. Journal of Physical Chemistry C, 116(37), 19781-19788. https://doi.org/10.1021/jp3044187

*To whom correspondence should be addressed

†Max-Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

‡Center of Smart Interfaces – Technische Universität Darmstadt, Petersenstr. 32, 64287 Darmstadt, Germany

Abstract

A simple classical force field, based only on Coulomb and Lennard-Jones potentials, was developed to describe the interaction of an ethanol molecule physisorbed on the α-alumina (0001) surface. A range of adsorption structures were calculated using density functional theory (DFT) and these results were used for the force field parameterisation. This system has a very inhomogeneous adsorption energy landscape and the importance of the choice of data set used for fitting the force field was investigated. It was found that a Lennard-Jones and Coulombic potential can describe the ethanol-alumina interaction in reasonable qualitative agreement with the DFT reference data provided that the data set was representative of both short and long range interactions and high and low energy configurations. Using a few distance-dependent adsorption energy curves at different surface sites gives the best compromise between computing time and accuracy of a Lennard-Jones based force field. This approach demonstrates a systematic way to test the quality of a force field and provides insight into how to improve upon the representability for a complex adsorption energy landscape.

Introduction

The interaction of soft matter with inorganic surfaces is of importance in many applications such as biomineralisation, composite materials and self-assembled monolayers. For protective polymeric coatings strong adhesion between soft matter and inorganic surfaces is essential. As a specific example polyurethane coatings are often applied to a surface as a mixture of dialcohols, trialcohols and dicyanates, which polymerise *in situ*. The interaction of the individual components with the surface will determine whether the mixture becomes segregated, which will affect the quality of the polymer coating, and how the final polymer bonds to the surface.

To understand the behaviour of such a system it is necessary to consider both the detailed chemical interaction at the surface, as well as the properties of the polymer or mixture. Classical molecular dynamics simulations can be used to understand the structure and dynamics of liquids on surfaces. For solid materials, Tersoff potentials,[1] Gaussian Approximation Potentials[2] or em-

bedded atom potentials[3] are applied, whereas in soft matter systems simple pairwise potentials in combination with a Coulomb interaction potential are frequently used. One of the most commonly employed pairwise potentials is the simple and computationally efficient Lennard-Jones (LJ) potential

$$V_{ij}^{\mathrm{LJ}} = 4\varepsilon_{ij}\left\{\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right\} \tag{1}$$

where ε_{ij} and σ_{ij} are adjustable parameters that determine the specific pairwise interaction between atom types i and j, separated by a distance r_{ij}. In this work, we apply an LJ potential in combination with Coulomb interactions to model the interaction between ethanol and the alumina. For such interface systems, the interaction parameters are typically taken from standard force fields, such as OPLS,[4,5] GROMOS[6–8] or AMBER.[9–11] However, these force fields were developed to describe bulk properties of liquids or solutions and it is unlikely that this approach would provide a qualitatively correct physical picture of a soft-hard matter interface. A better approach for developing a classical force field is to fit the parameters such that the results from the force field agree with the results of quantum chemistry or density functional theory (DFT) calculations.[12–22] Often, the interaction energy as a function of distance from the surface to the molecule at only a single or a few surface sites is considered. For relatively homogeneous surfaces, such as metallic surfaces, this might be a reasonable approximation. However, in the case of oxide surfaces, where the surface is very inhomogeneous, this inhomogeneity could strongly affect properties such as diffusion along the surface, local structure or adhesion strength. Thus, it is not *a priori* clear if force fields that are fitted to only a few adsorption states represent different adsorption states correctly. Here, we study the representability of force fields that are fitted to a subset of the adsorption energy landscape and investigate the optimal set of DFT data needed to obtain a representative force field.

This work thematically follows a previous study of a water molecule on ZnO(0001), which investigated the effect of fitting the force field parameters to a horizontal scan of the surface at constant height (*xy*-scan).[20] It was found that fitting to only a few adsorption configurations can bias the resulting force field and, therefore, the force field does not represent the full adsorption energy landscape. Here, we go a step further and compare the pros and cons of fitting to horizontal

surface scans versus fitting to z-dependent adsorption curves, which are often used in the literature to obtain classical force fields.[13–16,21,22] As a model system we take ethanol on the well-studied Al-terminated α-Al_2O_3(0001) surface, shown in Figure 1. The –OH group of the ethanol interacts via electrostatic interactions with the surface Al and O atoms and therefore the energy will vary considerably as the molecule moves across the solid surface. In total, we fit to 10 pair potentials (20 parameters) to three sets of adsorption energy landscape data, which in total includes 174 adsorption configurations (containing z-dependent and horizontal xy surface scans). The many adjustable parameters and the amount of DFT data makes a manual fit unfeasible and, therefore, we apply a genetic algorithm to optimise and automate the procedure, which has proven to be a useful tool in our previous work.[20]

This paper is organised as follows. First, in section Eq. (1), we outline the technical details of the DFT calculations and the classical force field. Next, we present DFT results for the interaction of the ethanol molecule with the surface in section Eq. (2). This is followed by a discussion of the fitting procedure, an evaluation of the representability of the force fields and validation in section Figure 3. The summary is presented in section Table 6.

Method

Density functional theory calculations were performed using the VASP code,[23,24] with a self-consistent van der Waals implementation[25,26] and PBE exchange.[27–29] The core electrons were described using projector augmented waves (PAW).[30,31] A planewave energy cutoff of 500 eV was used and all calculations used a Brillouin zone mesh equivalent to $4\times4\times1$ for the surface unit cell. Bulk alumina is hexagonal with an equilibrium lattice constant of a_0 =4.82 Å and $c/a = 2.73$. For the adsorption calculations we used a AlO_3Al-R slab, 18 atomic layers deep and 2×2 surface unit cells wide, which has lateral dimensions of $a = b = 9.64$ Å. Relaxation of the isolated molecule and isolated surface was stopped when the maximum force on any atom was less than 10 meV/Å. For all adsorption configurations the surface and molecule atoms were fixed. Partial charges were

calculated using the Bader analysis method.[32–34]

Classical calculations were performed using GROMACS 4.[35] The cutoff for the short range potentials was 13 Å and the electrostatic interactions were calculated using the Ewald method with a force and potential correction to avoid interactions between slab images in the z direction. The alumina surface was described by placing Al and O atoms in the positions of the relaxed isolated surface that were found from the DFT calculations. The classical simulations required a cell larger than the cutoff distance so the surface was multiplied by 3×3 along the a and b axis. The resulting simulation box was hexagonal with $a = b = 28.92$ Å and c =39.48 Å. Otherwise the molecule and surface atomic positions were identical to those used in the DFT calculations.

For the classical force field the interaction between the molecule and the surface was represented using electrostatic interactions and the well-known Lennard-Jones 12–6 potential with two parameters, σ_{ij} and ε_{ij}, per atom pair ij. The system had 7 atom types and 10 atom pairs, which resulted in the 20 different force field parameters listed in Table 4. Many force fields use the ionic formal charges, which are much higher than the actual charges present in the system. For the alumina surface, the currently developed force fields used the DFT Bader partial charges, presented in Table 1, which correspond to the charge contained between the charge density minima contours around the atoms. Each layer in the surface contained a formula unit of alumina with the atomic layers [Al_{upper}–3O–Al_{lower}], as shown in Figure 1(a). The data in Table 1 gives the average value of the three coplanar oxygen atoms. For the partial charges of the molecule the OPLS parameters were used, which are given in Table 2. This ensured that the the properties of liquid ethanol as predicted by the OPLS force field[5] were not altered.

The genetic algorithm used here has been described in detail in Ref.[20] The difference between the classical and the DFT energy landscape is $\delta E_j = E_{\mathrm{DFT}} - E_{\mathrm{class}}$ for a particular adsorption configuration j. The root mean square deviation (rms) of the energy difference for the ith parameter set is

$$\Delta_i = \sqrt{\frac{1}{M}\sum_{j=1}^{M} \delta E_j^2}, \tag{2}$$

Table 1: Partial charges of the isolated surface calculated using density functional theory. Al_{upper} (Al_{lower}) refers to the Al atom just above(below) the O atomic layer, as shown in Figure 1(a).

Layer	Al_{upper}	O	Al_{lower}
1	+2.42	−1.63	+2.49
2	+2.50	−1.66	+2.49
3	+2.49	−1.66	+2.49
4	+2.49	−1.66	+2.49
5	+2.49	−1.65	+2.49
6	+2.48	−1.57	+2.17

Table 2: Partial charges of ethanol used in the OPLS force field.

Atom	Partial charge
C_{OH}	+0.145
C_{H_3}	−0.180
H_C	+0.060
O_H	−0.683
H_O	+0.418

where M is the number of DFT data points used in the fitting procedure. The rms value is used to assess the accuracy of the force field.

Results and Discussion

Density functional theory

In a simulation of liquids on surfaces it may be necessary to use a classical potential that describes the vibrational and dynamical properties of the solid substrate. However, this work addresses only the interaction between the surface and the molecule. Furthermore, we restrict the configurations to non-dissociative adsorption structures of ethanol on alumina. In order to exclude any strain energy for surface relaxations, a series of unrelaxed configurations is used for the surface–molecule interaction. The surface atoms are fixed in the positions corresponding to that of the relaxed iso-

lated surface. In the relaxed isolated surface the top Al_{upper} atom sinks into the O layer to minimise the surface dipole, which can be seen in Figure 1. The molecule is translated rigidly in the x, y and z directions and various horizontal and vertical orientations are considered. This includes five different surface scans in the xy-plane with different heights from the surface and different molecular orientations and seven scans along the z direction at different surface sites. The various configurations are summarised in Table 3. For convenience x, y and z are given in fractional coordinates, which are in units of the simulation cell along the a, b and c axes.

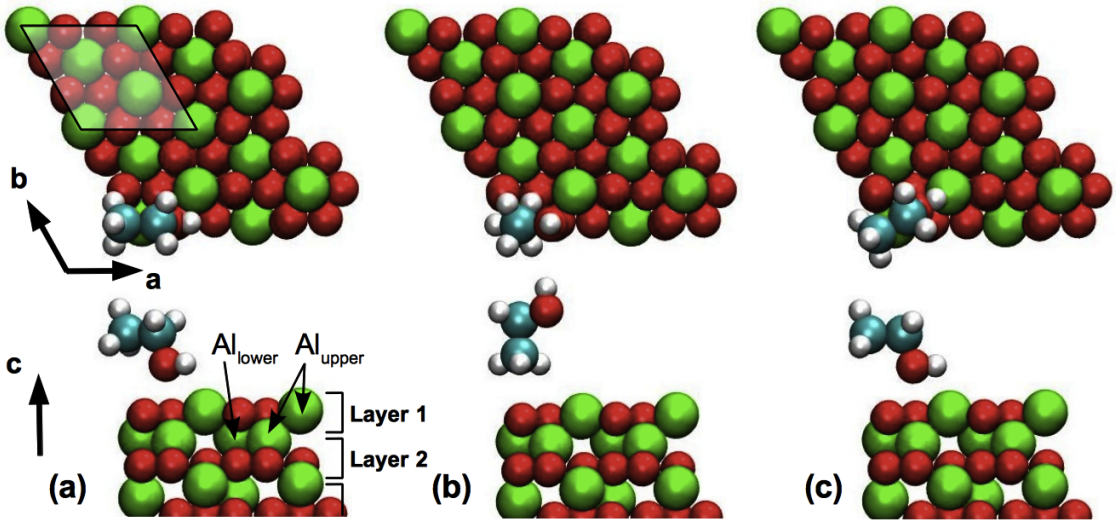

Figure 1: Three different orientations of ethanol on the alumina surface (a) $\theta = 0^o$, $\phi = 0^o$, corresponding to configurations xy-a, xy-b and xy-c (b) $\theta = 90^o$, corresponding to xy-d and (c) $\theta = 0^o$, $\phi = 30^o$, corresponding to xy-e. The rhombus in (a) shows the surface unit cell of alumina.

First we consider the xy-scans. In Figure 1(a) the molecule has the C–C bond parallel to the surface plane ($\theta = 0$) and the a direction ($\phi = 0$) and the –OH group points towards the surface. Three different z values for this configuration were used, namely, $z = 0.375$, $z = 0.380$ and $z = 0.390$, corresponding to the carbon atoms being a distance of 2.84 Å, 3.03 Å and $z = 3.43$ Å from the surface. In Figure 1(c) the orientation is similar except that the C–C bond makes an angle of $\phi = 30$ with the a-axis and in this case $z = 0.390$. In Figure 1(b) the molecule is in a vertical orientation with the C–C bond perpendicular to the surface plane $\theta = 0$, the –OH group pointing away from the surface and $z = 0.390$.

Table 3: Summary of the various scans used for the adsorption energy landscape. x, y and z are in fractional coordinates of the simulation cell and angles are in degrees.

	x	y	z	θ	ϕ
xy-a	–	–	0.375	0	0
xy-b	–	–	0.380	0	0
xy-c	–	–	0.390	0	0
xy-d	–	–	0.390	90	–
xy-e	–	–	0.390	0	30
z-a	0.00	0.00	–	0	0
z-b	0.20	0.10	–	0	0
z-c	0.20	0.30	–	0	0
z-d	0.25	0.25	–	0	0
z-e	0.20	0.00	–	0	0
z-f	0.00	0.00	–	90	–
z-g	0.00	0.00	–	0	30

Three of these scans are shown in Figure 2. Although the surface cell is hexagonal the scans are shown using orthogonal axis, where the x and y coordinates are the fractional coordinates of the surface simulation cell along the a and b directions. Figure 2(a) is for $z = 0.380$ with molecular orientation $\theta = 0$ and $\phi = 0$, and is mainly repulsive except at (0.2,0.1), which corresponds to the ethanol oxygen being close to the surface aluminium atom. Figure 2(b) is slightly further from the surface with z=0.390 and shows a mixture of attractive and repulsive sites. The shape of the landscape is similar to Figure 2(a) but shifted to lower energies. Figure 2(c) is at the same height as Figure 2(b) but with the molecule rotated by $\phi = 30^{\text{o}}$ in the xy-plane, as shown in Figure 1(c). In summary, the attractive regions correspond to configurations where the O_H is close to an Al atom and far from the O_s atoms. The xy-scans that are not shown are xy-a, which is entirely repulsive, and xy-d (shown in Figure 1(b)), whose adsorption energy only varies between -20.94 and -29.52 kJ/mol and is nearly featureless.

Next, we consider the seven different z-curves, which are shown in Figure 3. The most attractive site is at (0.20,0.10), as seen before in Figure 2, with a configuration of $\theta = 0^{\text{o}}$, $\phi = 0^{\text{o}}$ and with a minimum adsorption energy of around -90 kJ/mol at $z = 0.390$ (2.84 Å from the surface). The configuration with a vertical orientation of the molecule has the –OH group oriented away from

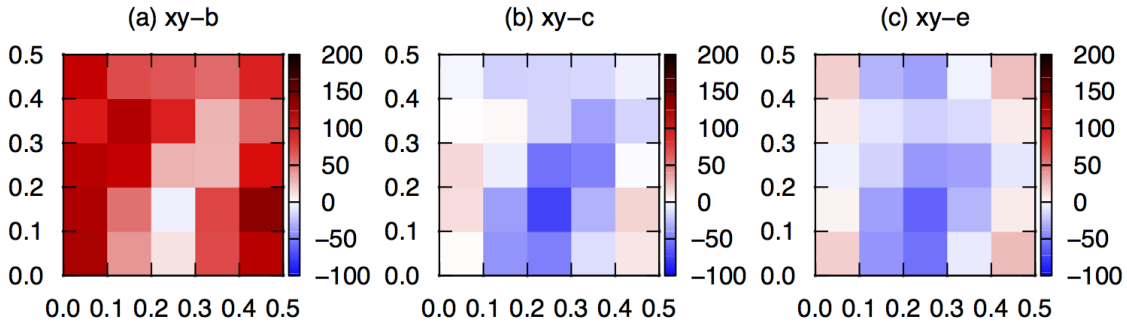

Figure 2: Horizontal scans of the interaction energy of ethanol on the alumina surface corresponding to (a) xy-b, (b) xy-c and (c) xy-e. Red corresponds to a repulsive interaction and blue is an attractive interaction. The x and y axes represent fractional coordinates along the a and b axes, respectively and the color bar range is in kJ/mol.

the surface and the curve is weaker and broader. The most weakly attractive site is at (0.0,0.0) with $\theta = 0^{o}$ and $\phi = 30^{o}$.

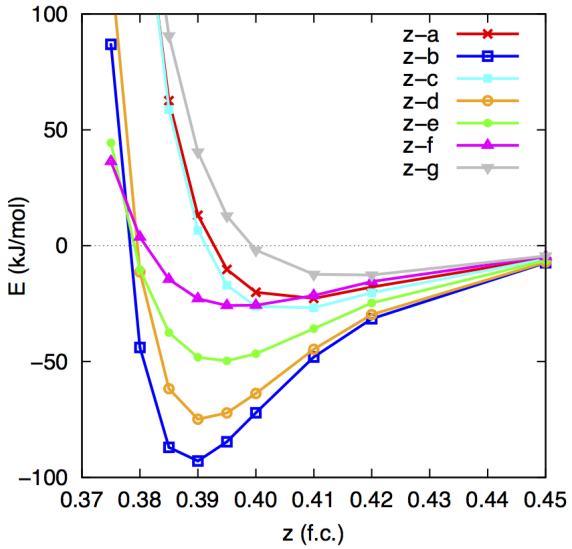

Figure 3: Vertical scan of the interaction energy of ethanol at different points on the alumina surface.

Force field optimisation

The density functional calculations in the previous subsection provide a good representation of the adsorption energy landscape. In this section we investigate whether the simple Lennard-Jones

and Coulomb force field can represent the DFT data and the importance of the choice of the data set used for optimisation. Additionally, we compare the performance of our present force fields with the results using a parameter set that was proposed by Youngs *et al* in a previous study of isopropanol on γ-alumina.[19]

The fitting procedure is challenging since there are many parameters and configurations involved and a compromise must be found between searching over a large parameter space and refining the parameters. For the genetic algorithm we used 64 parameter sets and a mutation rate of 3%. Sets 1 and 2 were run for 900 iterations and set 3 for 236 iterations. The reduced number of iterations for set 3 is because the large number of configurations is more computationally demanding than the other sets. The results for a particular dataset do not converge to a unique solution and several parameter sets can give similar rms values. For each dataset we report the parameter set with the lowest rms value. The lowest rms values were reached after 137, 93 and 117 iterations for sets 1, 2 and 3, respectively.

Since the electrostatics is a significant part of the interaction, we start by looking at the rms values obtained from the electrostatic interaction only. This gives rms values of 95 kJ/mol and 91 kJ/mol for the formal charges and DFT partial charges, respectively. By looking at the z-scans in Figure 4 it is clear that in some cases the electrostatics predict an unrealistic attractive interaction at short range, most notably in Figure 4(a). It is this short-range attraction that the repulsive part of the Lennard-Jones pair potential must balance. Next, we consider the force field by Youngs *et al*, which reproduces the Al-O distance and interaction energy for one configuration of a water molecule adsorbed on α-alumina.[19] The parameters for this force field are listed in Table 4. Using these parameters with the formal charges results in an rms of 290 kJ/mol and with partial charges the rms is 295 kJ/mol, which are both worse than using only electrostatics. The reason for this high rms is due to the inaccuracy of the close-range repulsion, which in Figure 4(b) swamps the attraction completely and in Figure 4(c) overestimates the adsorption energy.

To develop a more accurate force field it is necessary to optimise the Lennard-Jones pair potential parameters. In our previous paper for water on ZnO,[20] we optimised the parameters so that

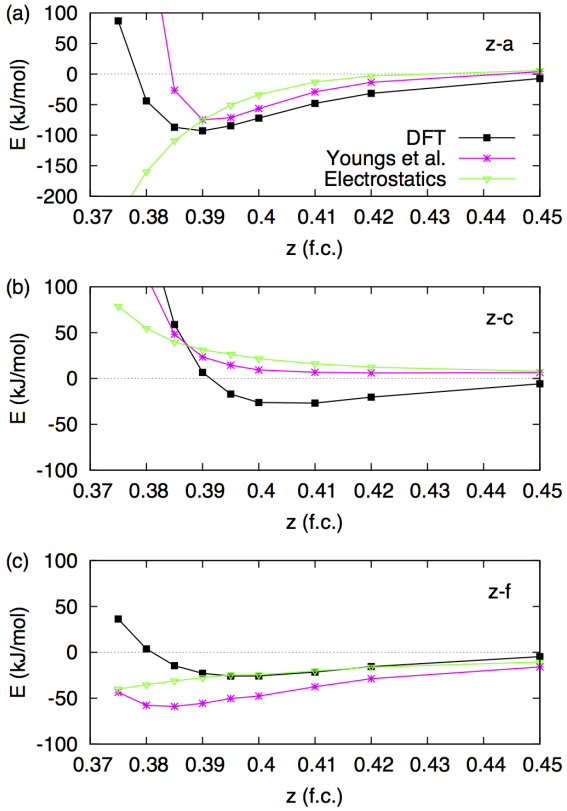

Figure 4: Comparison of DFT, electrostatics and classical energies using the parameter set from Youngs *et al.*[19] The three panels correspond to the three different configurations labelled in the graphs, corresponding to (a) z-b (b) z-c and (c) z-f.

the force field reproduced DFT results for an energy landscape in an *xy*-plane. The importance of fitting a wide range of adsorption configurations was highlighted. Here we take the work a step further by using a DFT data sets in *xy* and *z* and with with various configurations. The results obtained using the entire set and two subsets are compared. In total there are five scans of the adsorption energy in the *xy*-plane and *z*-scans for seven different configurations, which altogether gives 174 distinct data points. We have used the following three data sets for fitting (see Table 3):

Set 1 One *xy*-scan: xy-c (25 points).

Set 2 Three *z*-scans: z-b, z-c, z-f (27 points).

Set 3 Full data set (174 independent points).

The *xy*-scan in set 1, shown in Figure 2(b), was chosen because it has both repulsive and attractive

sites and contains the most strongly attractive site. The three z-scans of set 2 were chosen to represent both strongly and weakly attractive sites, as seen in Figure 3. The number of points in both of these sets is similar, which makes the comparison between these two optimised force fields fair.

Table 4: Force field parameters for the molecule–surface interaction. Units of σ and ε are nm and kJ/mol, respectively.

		Ref.[19]		Set 1		Set 2		Set 3	
Mol	Surf	σ	ε	σ	ε	σ	ε	σ	ε
C_{OH}	Al	0.30000	0.65	0.3563	0.6597	0.6245	0.0006	0.3375	1.0490
C_{H_3}	Al	0.30000	0.65	0.2135	0.2409	0.1304	0.3597	0.5079	0.0087
H_C	Al	0.30000	0.65	0.3255	0.9908	0.3117	0.6052	0.0239	0.1899
O_H	Al	0.26778	0.65	0.0196	0.4857	0.2684	0.0751	0.0130	0.5910
H_O	Al	–	–	0.1544	1.0340	0.2126	0.1280	0.3095	0.0409
C_{OH}	O_s	0.30000	0.65	0.3434	0.0014	0.3353	0.6278	0.1556	0.0088
C_{H_3}	O_s	0.30000	0.65	0.1699	0.8870	0.4448	0.2117	0.2019	0.3254
H_C	O_s	–	–	0.2670	0.4569	0.1244	0.1974	0.2120	0.5181
O_H	O_s	–	–	0.0981	0.8907	0.0114	0.6335	0.2295	0.6658
H_O	O_s	0.28000	0.65	0.1828	0.9992	0.2632	0.7019	0.2492	0.4860

All configurations with energy over 100 kJ/mol were weighted in the fitting procedure by 0.1 to account for the fact that the short-range behaviour of the Lennard-Jones potential is generally too repulsive. These very high energy configurations (in the order of tens of kT) are not likely to be accessed in molecular dynamics simulations and therefore the accuracy of these energies is less important. Nevertheless, the resulting force field should predict these configurations to be repulsive in order to avoid sampling unrealistic configurations. To compare the quality of the three parameter sets we use the unweighted rms value for the entire set. Table 5 shows the rms value for the weighted, fitted set (rms-fit), the total, unweighted rms value (rms-full), the rms values for short- and long-range and higher and lower energy configurations. Note that in set 1, all values are below 100 kJ/mol and, therefore, no weighting was required.

The resulting parameters are not "physical" in the sense that the σ values are not proportional to the ionic radii. In the context of fitting simple pair potentials to describe electronic structure data there is no reason to believe that the interactions should be simply represented by ions of a

particular size. These potentials encompass a range of multibody interactions and therefore the parameters should be allowed to vary freely so that they give the best fit to the available data. Nevertheless, the values of σ for the O_H–Al interaction in sets 1 and 3 are rather small and later we use stochastic dynamics simulations to test the accuracy of the force fields.

The rms values for the three fitted sets are shown in Table 5. For set 1, which is the fit to the adsorption energy landscape xy-c, we obtained an rms value for the fitted set of 21 kJ/mol and an rms value of 108 kJ/mol for the full set. This is an improvement over the parameter set used in Ref.[19] but is worse than the rms value obtained by using only the electrostatics without the pair potentials. The reason for this is that the electrostatics appear to describe well the long range energies but not the short range energies, as discussed previously. Sets 2 and 3 give an improved fitting, resulting in rms values for the full data set of 88 and 77 kJ/mol, respectively. Nevertheless, the rms difference between the DFT and classical energies is still quite large and the source of this inaccuracy will be analysed in more detail in the following text.

Table 5: Energies and rms differences of the classical energies vs. DFT energies for different charge and parameter sets in kJ/mol. The rms-fit is for the fitted set of data whereas rms-full is for the full dataset. The rms values are also given for the short-range ($z \leq 0.390$), long-range ($z > 0.390$), low energy (< 100 kJ/mol) and high energy (> 100 kJ/mol) configurations.

LJ	Charges	rms-fit	rms-full	$z \leq 0.390$	$z > 0.390$	$E_{DFT} < 100$	$E_{DFT} > 100$
Ref.[19]	Formal	–	289.63	–	–		
None	Formal	–	95.03	–	–	–	–
None	Partial	–	90.52	99.83	33.90	57.87	165.62
Set 1	Partial	20.96	107.70	120.03	21.36	83.09	173.93
Set 2	Partial	18.67	87.85	97.59	23.39	57.94	158.23
Set 3	Partial	40.85	76.79	84.71	28.53	39.08	152.47

First, we compare the results of the three parameter sets for the z-curves, shown in Figure 5. For set 1, which was fit to the xy-plane at $z = 0.390$, the energies at $z = 0.390$ are in excellent agreement with the DFT data and also agree reasonably well at longer distances. However, for short distances the curves diverge from the DFT data and can be very unphysical, as in Figure 5(a). Clearly, set 1 does not reproduce the z-dependence. This is likely to be true in our previous work on water on ZnO[20] where the fit to the data in the xy-scan was good but the z-dependence was not

tested. Unsurprisingly, set 2, which was fitted to these three curves, reproduces the energies rather well over the entire range, although it misses the weak attraction in Figure 5(b). Set 3, which was fitted to the entire DFT data set, does not give a good fit to the z-curves and in both Figure 5(a) and (c) it seriously underestimates the short range repulsion.

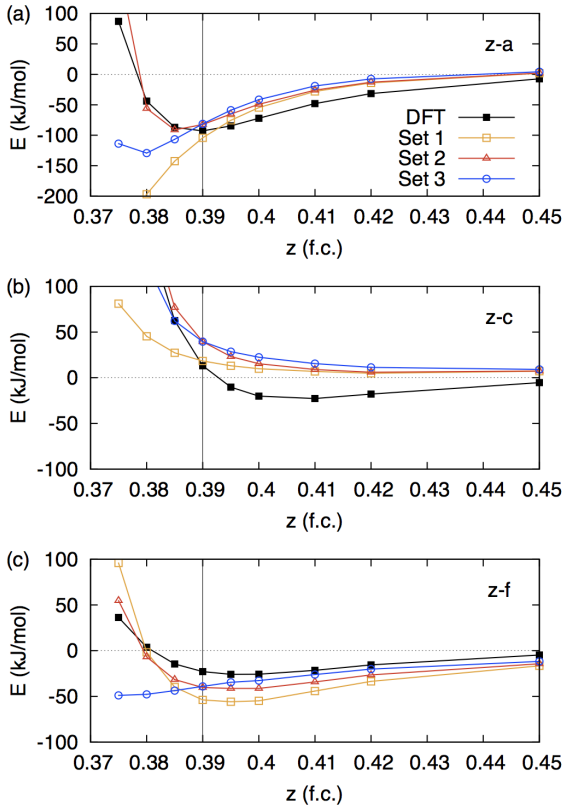

Figure 5: Comparison of DFT and classical energies for the z-scans (a) z-b (b) z-c and (c) z-f, using set 1 (xy-scan), set 2 (z-scan) and set 3 (full set).

Next, we compare how well the three parameter sets reproduce the xy-scans at $z = 0.380$ and $z = 0.390$, both with $\theta = 0^{\circ}$ and $\phi = 0^{\circ}$. Set 1 was fitted to xy-c, shown in Figure 6(e), and reproduces the DFT energies reasonably well (compare Figs. Figure 6(e) and (f)). However, this set does not reproduce well the xy-scan at $z = 0.380$ (compare Figs. Figure 6(a) and (b)), which is consistent with the results of the z-scans, where the energies at short distances are unphysical. Set 3 reproduces both energy landscapes better than set 1 (compare Figs. Figure 6(a) with (d) and (e) with (h)) but wrongly predicts some of the short-range repulsive configurations to be attractive.

Set 2 gives the best agreement with the DFT data for both *xy*-scans. This is surprising given that the rms value for set 3 is lower than for set 2 and to understand this we must consider the full set of data.

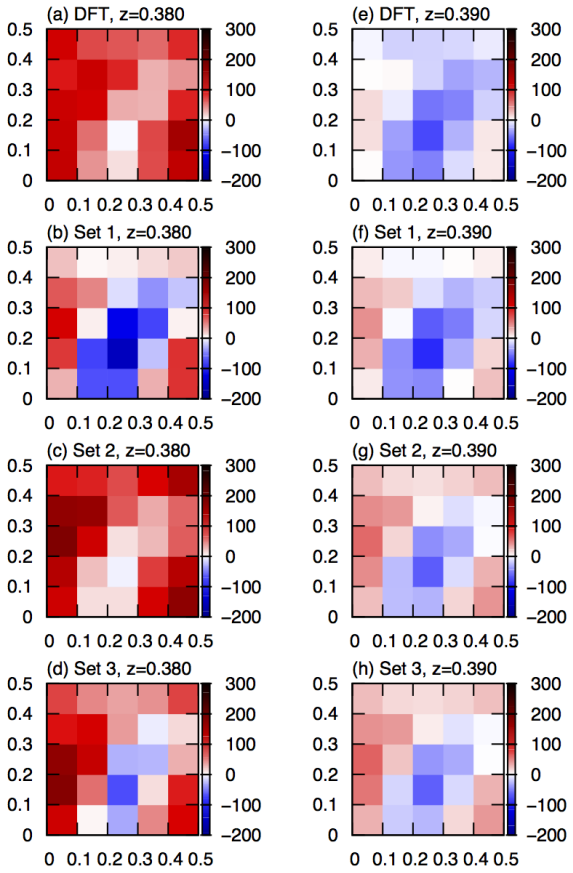

Figure 6: Adsorption energies for dataset xy-b using (a) DFT, (b) set 1, (c) set 2, (d) set 3 and for dataset xy-c using (e) DFT (f) set 1, (g) set 2 and (h) set 3. The x and y axes are in fractional coordinates and the color bar range is in kJ/mol.

The full data set for all parameter sets is shown in Figure 7. As mentioned before the electrostatics alone give a low rms value of 91 kJ/mol and the DFT adsorption energies in the range -100 to $+100$ kJ/mol are reproduced reasonably well. However, in Figure 7(a) one can clearly see that the problem with the electrostatics is in the high DFT energy range, which corresponds to the repulsive regime close to the surface. In addition, there are some outlying points not shown on the graph with classical energies below -200 kJ/mol but higher DFT energies. For these higher energy configurations electrostatics alone predict much too low energies as we have already seen before

in Figure 4. The results from set 1 are shown in Figure 7(b) and, similar to the electrostatics, it

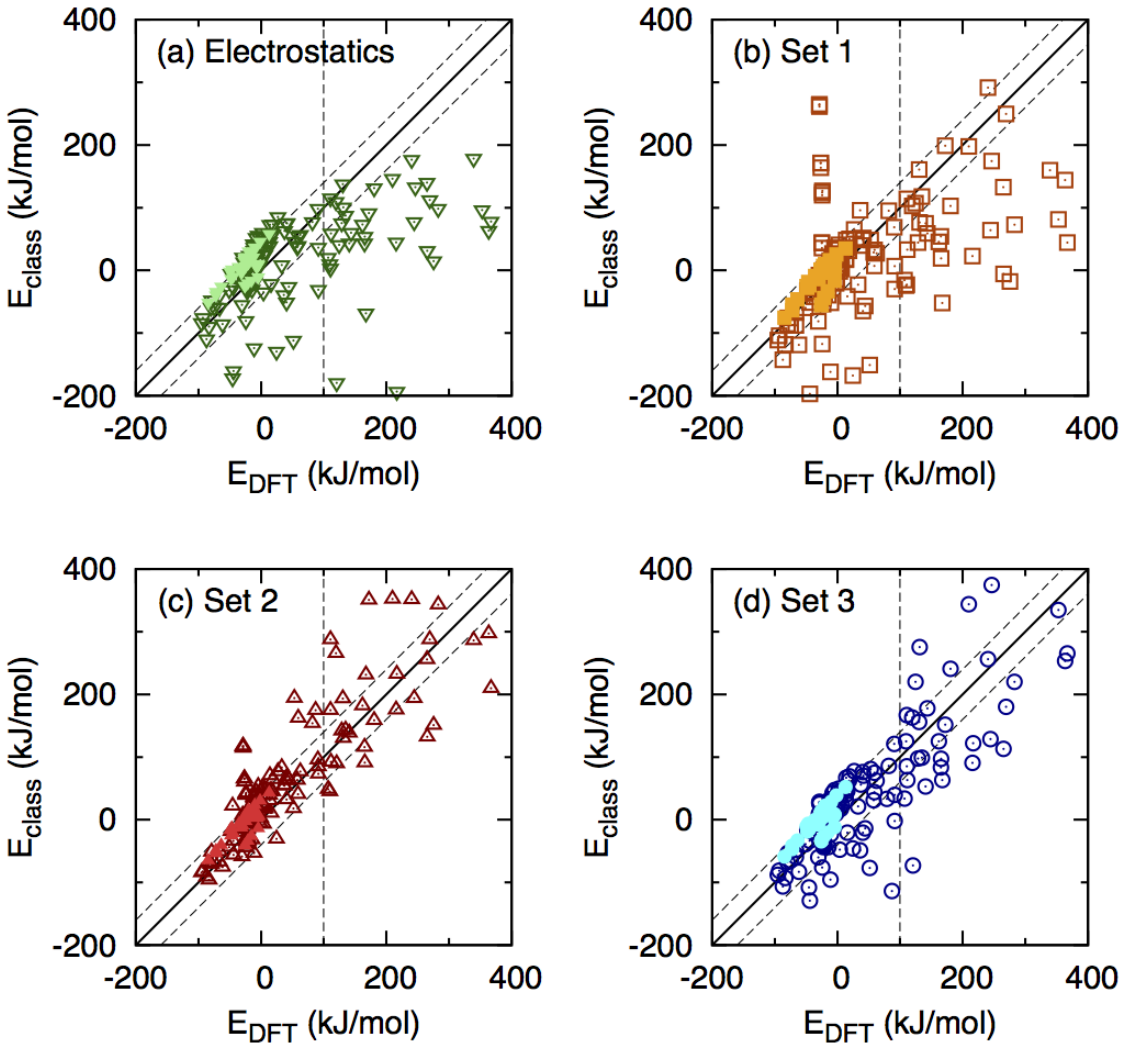

Figure 7: The classical energies shown against the DFT energies for all the data points using (a) electrostatics only, (b) set 1 (*xy*-scan), (c) set 2 (*z*-scans) and (d) set 3 (full set). The long range configurations are presented as filled symbols and the short range configurations as open symbols. The dashed lines are guides for the eye and correspond to an energy difference of ± 40 kJ/mol.

seriously underestimates the energy of many high energy structures. As mentioned previously, this would lead to incorrect sampling of high energy states and unrealistic simulations. Set 2 and set 3 have corrected this behaviour by giving a more accurate description of the repulsive configurations. Sets 1 and 2 have 5 and 3 outlying points, respectively, with very high energy classical energies that are not shown in the graphs. However, these points have DFT energies above 100 kJ/mol, except

for one point in set 2 that has $E_{\mathrm{DFT}} = 78.9$ kJ/mol, and, therefore, these configurations have a low sampling probability in molecular dynamics simulations at ambient temperature. In the regime below 100 kJ/mol (to the left of the dashed line in Figure 7), set 3 has a lower rms value than set 2, as seen in Table 5. It can also be seen in Figure 7 that set 3 tends to underestimate the energies whereas set 2 tends to overestimate them.

For all three sets the biggest discrepancy between the classical and DFT energies originates from the regime where the molecule is in close proximity to the surface, as seen in Figure 7. In Table 5 we reported separate rms values for configurations in two different regimes. The first regime corresponds to configurations close to the surface ($z \leq 0.390$) and the second regime to configurations at distances $z > 0.390$. We have seen that sets 2 and 3 reproduce these short-range configurations better than set 1 and the electrostatics.

Although the overall rms for set 2 is slightly higher than the one for set 3, as shown in Tab. Table 5, this should not be the only criteria for judging the representability of the force field. For example, set 2 was fitted to three *z*-scans but also obtained *xy*-scans in qualitative agreement with the DFT results, whereas set 3 was fitted to all the data but failed to qualitatively predict the xy-c, z-b and z-f data. Furthermore, the targeted optimisation of set 2 in comparison to the full set is computationally less demanding.

To test the validity of the force fields we ran stochastic dynamics simulations of an ethanol molecule on the surface. Two different starting configurations were used for each force field and the simulations were run at 300 K for 100 ps. In all cases the ethanol finds a stable configuration within 3 ps, with the ethanol oxygen close to a surface aluminum, and remains there for the duration of the simulation. The structural properties were averaged over the time that the molecule remained in this minimum energy well. The different force fields give quantitatively different structural properties. The distance between the surface aluminium atom and the ethanol oxygen, $z_{\mathrm{O_H-Al}}$ is presented in Table 6. Clearly for sets 1 and 3, $z_{\mathrm{O_H-Al}}$ is much too small compared to the DFT data, which is due to the small value of $\sigma_{\mathrm{O_H-Al}}$ for these sets. This is connected to the fact that set 1 does not include any short-range configurations so the short-range repulsion is not taken into

account in the fitting. Although set 3 includes the short-range configurations the data set has many mid-range configurations i.e. at $z = 0.390$, which biases the fitting procedure. Therefore, sets 1 and 3 do not represent the energy landscape well at close distances, consistent with graphs in Figure 5, Figure 6 and Figure 7. The distance of $z_{O_H-Al} = 0.183$ nm, for set 2 agrees well with DFT data for methanol on alumina[38] and is slightly better than the value found using the force field of Ref.[19] A stochastic dynamics simulation using only the electrostatics proved to be unstable.

Table 6: Comparison of the bond length between the oxygen ethanol and the Al_{upper} atom on the surface predicted by the three force fields, the force field used by Youngs[19] and *ab initio* calculations of methanol on alumina.[38]

	z_{O_H-Al} (nm)
Set 1	0.100
Set 2	0.183
Set 3	0.117
Ref.[19]	0.221
Ref.[38]	0.193

Conclusions

In this paper we have presented a strategy for optimising interface force fields for ethanol on an alumina surface using a genetic algorithm. Density functional theory including van der Waals interactions was used to calculate the binding energies of a series of configurations of ethanol on alumina and the partial charges of the alumina surface. Three datasets were used for the fitting procedure: 1) an *xy*-landscape 2) three *z*-dependent scans of the adsorption energy at different surface sites and 3) a larger dataset contains five *xy* landscapes and seven *z*-dependent adsorption energy curves. The force field consists of electrostatic interactions and a non-bonded Lennard-Jones 12–6 pair potential.

An appropriate reference dataset is essential for obtaining a representative force field. A dataset only reproduces the entire adsorption energy landscape well if the dataset contains both short-

and long-range and low and high energy configurations. Fitting to a single *xy* landscape resulted in a force field that did not reproduce the *z*-dependence, whereas fitting to a few *z*-curves on both repulsive and attractive sites gave qualitatively good agreement for the *xy* energy landscapes. Fitting to the entire dataset is computationally costly and resulted in a poorer quality force field than the force field fitted using only the three *z*-curves. This is attributed to the fact that the dataset was not distributed evenly across the potential energy landscape and biased the fitting towards mid-distance configurations.

Despite the simplicity of this force field, we have shown that by optimising the force field parameters it is possible to obtain reasonable agreement with DFT data and, hence, to avoid sampling of unrealistic configurations in molecular dynamics simulations. Unfortunately, the validity of interface force fields in the literature has rarely been thoroughly checked. This work illustrates the difficulties involved in optimising a force field that describes accurately an inhomogeneous adsorption energy landscape. Nevertheless, we have clearly demonstrated a systematic approach for checking and improving upon the representability of such a force field.

Clearly, this simple force field does have limitations, especially in the short range regime. This force field development could be improved by changing the type of pair potential used. The Lennard-Jones 12-6 potential is known to be too repulsive at short distances and the use of a different pair potential form, such as the Morse potential,[21] could give better agreement in the short-range regime. A further improvement would be to include polarisability in the force field.

Acknowledgments

The authors thank Konstantin Koschke and Sebastian Fritsch for critical reading of the manuscript. This work was supported by the Multiscale Modelling Initiative of the Max Planck Society and the DFG project SPP 1369.

References

(1) J. Tersoff. Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Phys. Rev. B*, 39:5566, 1989.

(2) Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.*, 104:136403, 2010.

(3) M. S. Daw and M. I. Baskes. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Phys. Rev. B*, 29:6443–6453, 1984.

(4) W. L. Jorgensen and J. Tirado-Rives. The OPLS Force Field for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.*, 110:1657, 1988.

(5) W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.*, 118:11225, 1996.

(6) J. Hermans, H. J. C. Berendsen, W. F. van Gunsteren, and J. P. M. Postma. A consistent empirical potential for water–protein interactions. *Biopolymers*, 23:1513, 1984.

(7) C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, 25:1656–1676, 2004.

(8) L. D. Schuler, X. Daura, and W. F. J. van Gunsteren. An Improved GROMOS96 Force Field for Aliphatic Hydrocarbons in the Condensed Phase. *J. Comput. Chem.*, 22:1205, 2001.

(9) P. K. Weiner and P. A. Kollman. AMBER: Assisted Model Building with Energy Refinement. A General Program for Modeling Molecules and Their Interactions. *J. Comput. Chem.*, 2:287, 1981.

(10) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, 117:5179, 1995.

(11) D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, and P. A. Kollman. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.*, 91:1, 1995.

(12) A. V. Bandura and J. D. Kubicki. Derivation of Force Field Parameters for TiO_2-H_2O Systems from ab Initio Calculations. *J. Phys. Chem. B*, 107:11072, 2003.

(13) M. I. McCarthy, G. K. Schenter, C. A. Scamehorn, and J. B. Nicholas. *J. Phys. Chem.*, 100:16989–16995, 1996.

(14) Oleg Borodin, Grant D. Smith, Rajdip Bandyopadhyaya, and Oleksiy Byutner. Molecular dynamics study of the influence of solid interfaces on poly(ethylene oxide) structure and dynamics. *Macromolecules*, 36:7873, 2003.

(15) Pim Schravendijk, Nico F. A. van der Vegt, Luigi Delle Site, and Kurt Kremer. Dual-Scale Modeling of Benzene Adsorption onto Ni(111) and Au(111) Surfaces in Explicit Water. *Chem. Phys. Chem.*, 6:1866, 2005.

(16) Daniel J. Cole, Mike C. Payne, Gábor Csányi, S. Mark Spearing, and Lucio Colombi Ciacchi. Development of a classical force field for the oxidized Si surface: Application to hydrophilic wafer bonding. *J. Chem. Phys.*, 127:204704, 2007.

(17) Luca M. Ghiringhelli and Luigi Delle Site. Phenylalanine near Inorganic Surfaces: Conformational Statistics vs. Specific Chemistry. *J. Am. Chem. Soc.*, 130:2634, 2008.

(18) F. Iori, R. Di Felice, E. Molinari, and S. Corni. GoIP: An Atomistic Force-Field to Describe the Interaction of Proteins With Au(111) Surfaces in Water. *J. Comput. Chem.*, 30:1465, 2008.

(19) T. G. A. Youngs, D. Wever, L. F. Gladden, and C. Hardacre. Liquid Structure and Dynamics of Aqueous Isopropanol over γ-Alumina. *J. Phys. Chem. C*, 113:21342, 2009.

(20) Claudia R. Herbers, Karen Johnston, and Nico F. A. van der Vegt. Modelling surface interactions – an automated quantum-classical approach using a genetic algorithm. *Phys. Chem. Chem. Phys.*, 13:10577, 2011.

(21) Karen Johnston and Vagelis Harmandaris. Properties of Benzene Confined between Two Au(111) Surfaces Using a Combined Density Functional Theory and Classical Molecular Dynamics Approach. *J. Phys. Chem. C*, 115:14707, 2011.

(22) J. Schneider and L. Colombi Ciacchi. *J. Chem. Theory Comput.*, 7:473–484, 2011.

(23) G. Kresse and J. Hafner. Ab initio molecular dynamics for open-shell transition metals. *Phys. Rev. B*, 48:13115, 1993.

(24) G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mat. Sci.*, 6:15, 1996.

(25) M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. Erratum: Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.*, 95:109902(E), 2005.

(26) Andris Gulans, Martti J. Puska, and Risto M. Nieminen. Linear-scaling self-consistent implementation of the van der Waals density functional. *Phys. Rev. B*, 79:201105(R), 2009.

(27) John. P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, 77:3865, 1996.

(28) J.P. Perdew, K. Burke, and M. Ernzerhof. Erratum: Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, 78:1396(E), 1997.

(29) J. P. Perdew, K. Burke, A. Zupan, and M. Ernzerhof. Nonlocality of the density functional for exchange and correlation: Physical origins and chemical consequences. *J. Chem. Phys.*, 108:1522, 1998.

(30) P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953, 1994.

(31) G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758, 1999.

(32) G. Henkelman, A. Arnaldsson, and H. Jonsson. A fast and robust algorithm for Bader decomposition of charge density. *Comput. Mat. Sci.*, 36:254–360, 2006.

(33) E. Sanville, S. D. Kenny, R. Smith, and G. Henkelman. An improved grid-based algorithm for Bader charge allocation. *J. Comput. Chem.*, 28:899–908, 2007.

(34) W. Tang, E. Sanville, and G. Henkelman. A grid-based Bader analysis algorithm without lattice bias. *J. Phys.: Condens. Matter*, 21:084204, 2009.

(35) Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, March 2008.

(36) Vladimir Shapovalov and Thanh N. Truong. Ab Initio Study of Water Adsorption on α-Al_2O_3 (0001) Crystal Surface. *J. Phys. Chem. B*, 104:9859, 2000.

(37) E. M. Fernandez, R. I. Eglitis, G. Borstel, and L. C. Balbas. Ab initio calculations of H_2O and O_2 adsorption on Al_2O_3 substrates. *Comp. Mat. Sci.*, 39:587, 2007.

(38) O. Borck and E. Schröder. First-principles study of the adsorption of methanol at the α-Al_2O_3(0001) surface. *J. Phys.: Condens. Matter*, 18:1, 2006.

Graphical TOC Entry