Supporting Information: Unraveling the Pathways of Tribochemical Reactions Involving the ZDDP Lubricant Additive by Machine-Learning-Informed Molecular Dynamics

Enrico Pedretti, Francesca Benini, Giovanni Gravili, and Maria Clelia Righi*

Department of Physics and Astronomy, University of Bologna, Viale Berti Pichat 6/2, 40127, Bologna, Italy

E-mail: clelia.righi@unibo.it

Testing of the Machine Learning Potentials

Models details

We report the details for the three MLP models:

- **MACE**: for the model trained from scratch, we used a 2-layers message passing model with 128 channels and `max_L=1` (128x0e+128x1o), a cutoff of 4 Å, 8 radial basis, and body-order of 4 (`correlation=3`), enabling the "Agnesi" distance transform.
- **MACE-FT**: we fine-tuned the pretrained MACE-MATPES-PBE-0 model using the multi-head replay mode (from a replay dataset containing only structures with combi-

nations of elements in our dataset). The model size is analogous to the MACE model trained from scratch, the main difference being the cutoff, equal to 6 Å.

- **DP**: we used the `se_e2_a` descriptor, with a cutoff of 6 Å (`rcut_smth=5`), [25,50,100] neurons in the descriptor net (`axis_neuron=16`), and [120,120,120] neurons in the fitting net.

Parity plots

Fig. S1 contains the parity plots of predicted forces against DFT forces for the MACE model for all systems in the dataset. For each of them, mean absolute error (MAE) and root mean square error (RMSE) are reported, and the inset contains a picture of a representative structure of the system. The data points are presented as a two-dimensional histogram, where each bin corresponds to a force interval of 20 meV/Å, using a cold-to-hot color scheme to represent the density of data points. Ideally, all points should align with the diagonal line, corresponding to perfect agreement between MLP predictions and reference forces. For most of the systems the alignment is excellent, with the exception of systems containing Fe slabs. This is not surprising, mainly due to the presence of defects and steps, either manually included in the geometries or resulting from the extreme temperature and pressures at which part of the dataset configurations were sampled. Since Fe is magnetic, unstable geometries can lead to local spin rearrangements, resulting in multiple near-degenerate local minima. This can already be challenging at the DFT level, as the self-consistent cycle may converge to different magnetic states depending on the initial magnetic moments or wavefunction initialization. Consequently, some configurations in the dataset may correspond to metastable magnetic solutions rather than the true ground state for a given geometry, introducing noise into the training data. If number of these outliers is limited (such as in this case), the impact on the overall accuracy is minor. On the other hand, because MLPs tend to produce smooth potential energy surfaces as functions of atomic positions, they may struggle to accurately represent abrupt changes in energy associated with spin transitions or

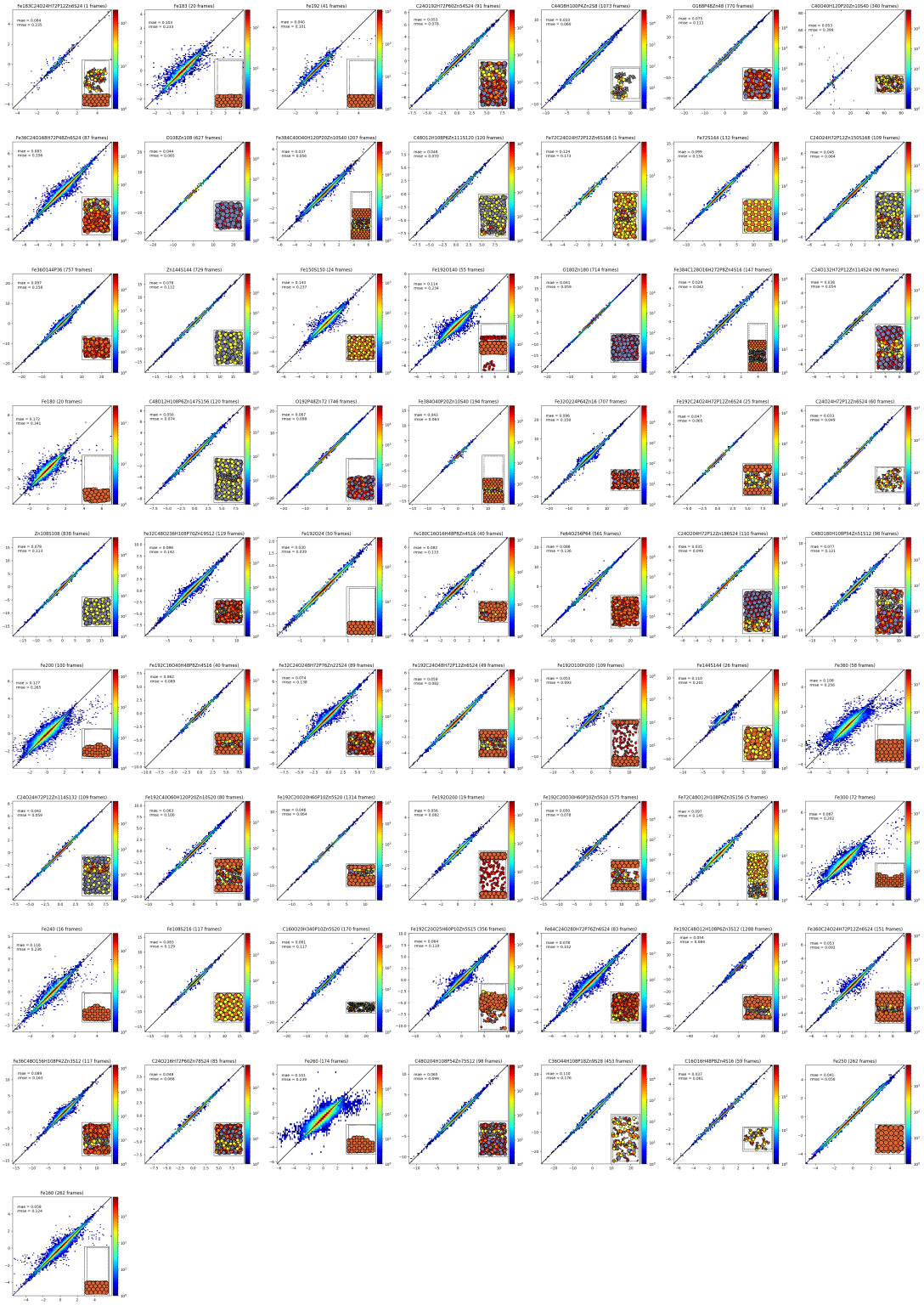

Figure S1: Parity plots for MACE model, comparing forces predicted by MLP (y axis) against reference DFT forces (x axis). The title of each plot states the elements and the number of frames (configurations) contained in the corresponding system. Data points are represented as 2D histograms (heatmaps) with the color bar on the side indicating the density of points.

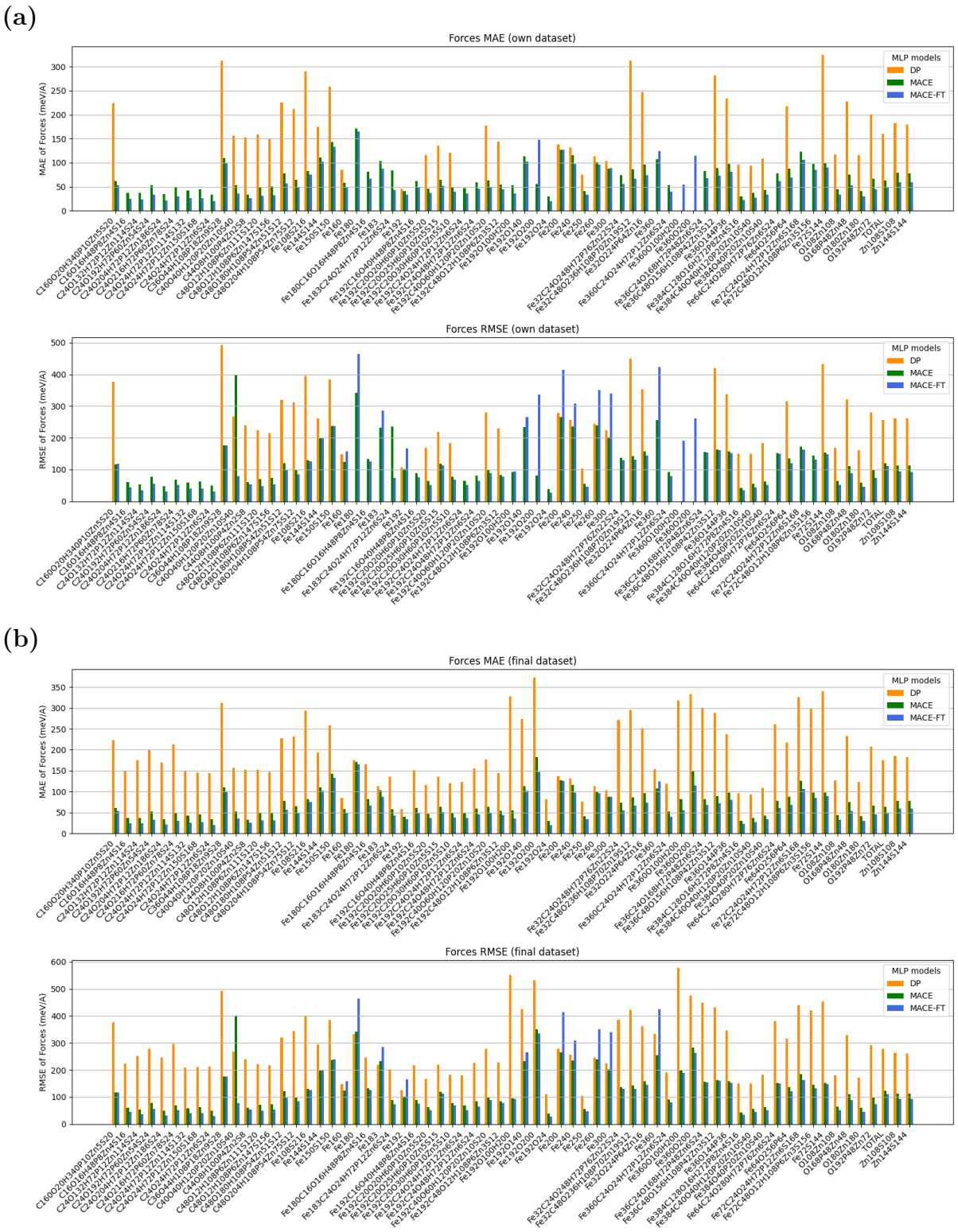

Figure S2: Comparison of force accuracy (MAE and RMSE) of DP, MACE and MACE-FT models on the dataset they were trained on (a) and on the final dataset (b), which coincides with MACE-FT dataset. The labels for each system are consistent with those in S1. In (a), missing columns corresponds to systems that were not present in the dataset used to train the corresponding model. The "TOTAL" entry corresponds to the average on all dataset.

reconfigurations. To assess the impact of these outliers on the MLP models, we performed a series of tests, removing increasing amounts of outliers from the training set and repeating the training procedure. This resulted in an improvement of just a few meV/Å in the force RMSE for most systems, at the cost of excluding training configurations that, despite the "problematic" local environments, may contribute to a better coverage of the configuration space. For this reason, the training for the production models was performed on the full dataset without removing outliers. In fact, even considering these outliers in the parity plots, the MAE and RMSE of the dataset falls well within what is generally considered to be accurate enough for a machine learning potential given the wide range of temperatures and pressures.

While we included all the parity plots just for the MACE model, we report the MAE and RMSE values on each system for all three models (DP, MACE and MACE-FT) in Fig. S2, both on the dataset used for the training of each model (which is smaller for DP and MACE compared to MACE-FT), and on the final dataset used to train the MACE-FT model. As expected, both MACE models are significantly more accurate than DP on all systems, which can be ascribed to the more advanced (and expensive) architecture. MACE-FT reaches the overall best accuracy. For a few systems, however, in particular those containing Fe, larger RMSE values are observed for MACE-FT (while still providing lower MAE). From the parity plots (not shown) a larger spread of outliers seems to be present. This is related to the point raised above: as spin local minima tend to be filtered out by the training, this effect is likely to be enhanced in finetuning, which is only a refinement of a pre-trained potential, while for the MACE model trained from scratch on our dataset overfitting of such outliers can occur.

Dissociation and isomerization energies

In Fig. S3, configurations used to test the MLPs against *ab initio* calculations are reported, including the three dissociation configurations (a-c) and isomerizations (d,e). For the dissociations, energy differences are calculated as the difference between the sum of dissociated

subsystems, and the sum of energies of the molecule in vacuum and a number of isolated Fe slab to match the number of subsystems.

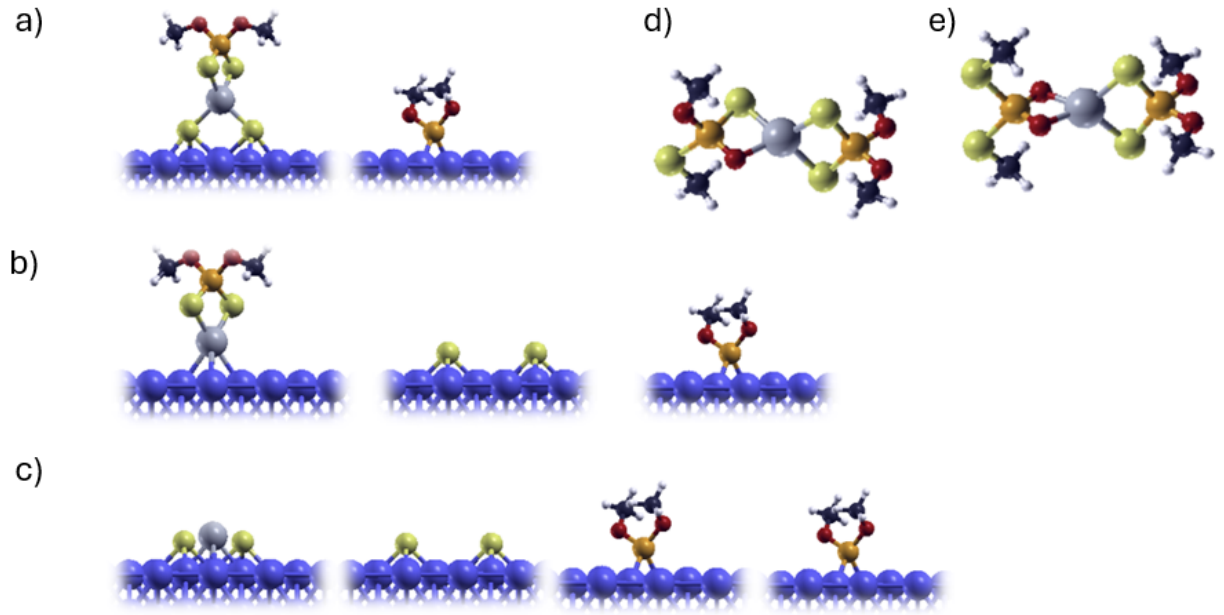

Figure S3: Configurations for dissociation 1-3 (a-c) and isomerization LI1 and LI2 (d, e) employed to test the MLPs on energy predictions.

NEB calculations of reaction pathways

The isomerization NEB reported in the main text for LI2-ZDDP, corresponding to Zn-O bond opening, forming a reactive P=O intermediate, and rotation of the phosphate group to form a Zn-S bond, was used to test MLPs accuracy, as visible in Fig. S4. As expected, both MACE models provide an excellent agreement with DFT, very accurately capturing the whole reaction path. The DP model instead is less accurate, but it still reproduces correctly the transition state with the same energy barrier.

Furthermore, meaningful reaction events observed during the MD simulations with MLPs were reproduced with DFT, calculating reaction barriers by means of DFT calculations, as reported in Fig. S5. Also in this case, the three MLP models were tested by recalculating statically the NEB images from DFT. All networks tend to capture quite accurately the

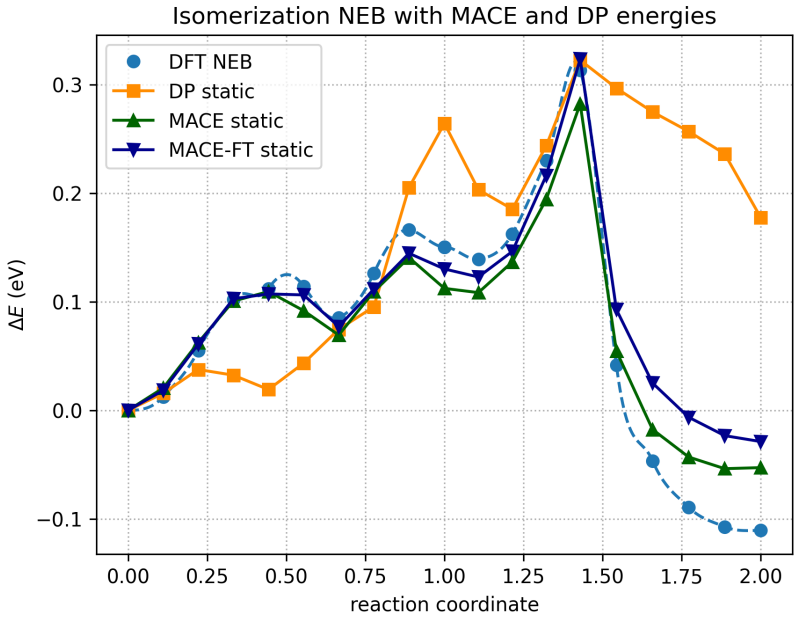

Figure S4: Testing of MLP models accuracy on the isomerization pathway reported in the main text, by recalculating the static images obtained through the full NEB calculation with DFT.

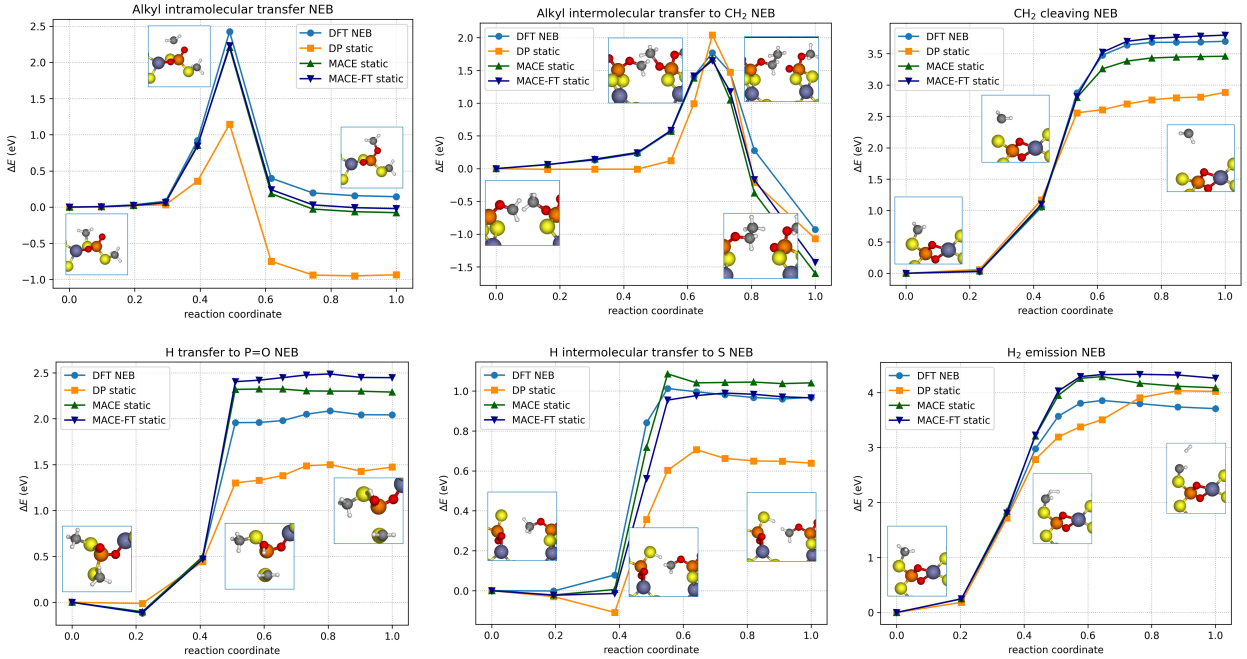

Figure S5: Additional NEB calculations for some significative reactive events observed in the machine learning MD simulations commented in the main text. Full NEB calculations were peformed with DFT, and static images were re-calculated with MLPs for testing their accuracy. The graph titles indicate the reaction, and inset images show the atomistic process.

reactions, with again MACE outperforming DP, which still shows a reasonable agreement in most cases. These tests serve as a confirmation of the validity of the events observed in MD simulations with MLPs.

Generation of the oxidized Fe(210) slabs

The oxidized Fe(210) surface was obtained through a MD simulation (performed with LAMMPS using the MACE-FT model) of spontaneous passivation by gaseous oxygen and water molecules. In the initial geometry, 1800 O_2 and 600 H_2O molecules were randomly placed using Packmol within a Fe(210)-Fe(210) interface (cell and slab sizes are reported in the main text). The outermost layers of the two slab were kept fixed, and the rest of the system was simulated in the NVT ensemble, controlling temperature with a single Nosé-Hoover thermostat for both slabs and the molecules inbetween, as no sliding is involved. The simulation lasted 1 ns, comprising an initial temperature ramp from 300K to 800K in 800ps, and a final cooling ramp from 800K to 100K in 200 ps. As visible in Fig. S6, after just 20 ps (at 300K) oxygen molecules attack the Fe surface, penetrating deeply and "extracting" Fe atoms which are vertically displaced from their original position to form an oxidized ∼400 K) most of the oxygen has reacted with Fe, having now formed a thick oxide layer (with surface z coordinate visbly altered from the intial one). Water molecules are mostly adsorbed on the surface, and some of them underwent dissociation, forming terminal Fe-OH groups. As temperature increases up to 800 K, the oxide layer stays in place and is slowly annealed to better accomodate some of the initial local strains, while part of water molecules are desorbed, going back to gas phase. Finally, at the end of the cooling ramp to 100K, water molecules re-adsorb again on both surfaces, and the oxide layer relaxes to its final structure.

The final geometry was used for the MD simulations with ZDDP after removing the physisorbed water layer, leaving only chemisorbed water and Fe-OH terminations.

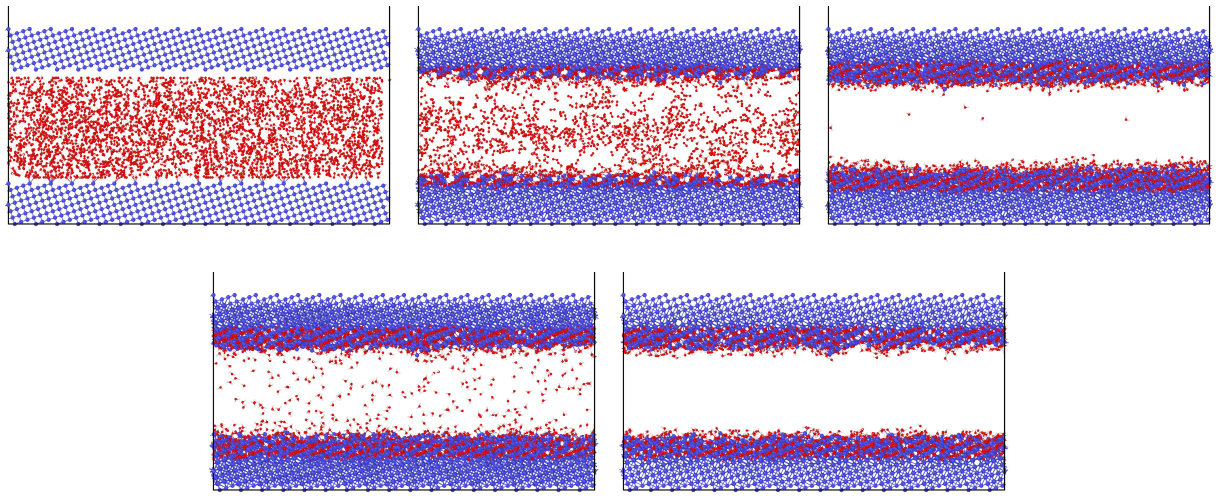

Figure S6: Steps of the MD simulation to obtain a spontaneously oxidized Fe surface. From top left corner to bottom right: initial configuration; snapshot at 20 ps (∼300K) showing already partial oxidation; snapshot at 200 ps (∼400K) where oxidation is complete (all O_2 has been consumed) and water is mostly adsorbed; snapshot at 800 ps (800 K) where the oxide layer is annealed and some water molecules desorb due to the high temperature; and the final snapshot at 1 ns (100K), where the oxide layer stabilizes and water adsorbs again on the surface.