Deringer, Volker L.; Caro, Miguel A.; Jana, Richard; Aarva, Anja; Elliott, Stephen R.; Laurila, Tomi; Csányi, Gábor; Pastewka, Lars

Computational Surface Chemistry of Tetrahedral Amorphous Carbon by Combining Machine Learning and Density Functional Theory

Published in:
Chemistry of Materials

DOI:
10.1021/acs.chemmater.8b02410

Published: 01/01/2018

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Deringer, V. L., Caro, M. A., Jana, R., Aarva, A., Elliott, S. R., Laurila, T., Csányi, G., & Pastewka, L. (2018). Computational Surface Chemistry of Tetrahedral Amorphous Carbon by Combining Machine Learning and Density Functional Theory. *Chemistry of Materials*, *30*(21), 7438–7445. https://doi.org/10.1021/acs.chemmater.8b02410

Computational Surface Chemistry of Tetrahedral Amorphous Carbon by Combining Machine Learning and DFT

Volker L. Deringer,[†,‡,*] Miguel A. Caro,[§,∥] Richard Jana,[#] Anja Aarva,[§] Stephen R. Elliott,[‡] Tomi Laurila,[§] Gábor Csányi,[†] and Lars Pastewka[#]

† Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, United Kingdom
‡ Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom
§ Department of Electrical Engineering and Automation, Aalto University, Espoo 02150, Finland
∥ Department of Applied Physics, Aalto University, Espoo 02150, Finland
Department of Microsystems Engineering, University of Freiburg, 79110 Freiburg, Germany

ABSTRACT: Tetrahedral amorphous carbon (*ta*-C) is widely used for coatings due to its superior mechanical properties and has been suggested as an electrode material for detecting biomolecules. Despite extensive research, however, the complex atomic-scale structures and chemical reactivity of *ta*-C surfaces are incompletely understood. Here, we combine machine learning, density-functional tight-binding, and density-functional theory simulations to shed new light on this long-standing problem. We make atomistic models of *ta*-C surfaces, characterize them by local structural fingerprints, and provide a library of structures at different system sizes. We then move beyond the pure element and exemplify how chemical reactivity (hydrogenation and oxidation) can be modeled at the surfaces. Our work opens up new perspectives for modeling the surfaces and interfaces of amorphous solids, which will advance studies of *ta*-C and other functional materials.

INTRODUCTION

Understanding the structures, properties, and reactivity of material surfaces is a challenging task requiring a wide array of experimental and computational tools. Today, atomistic simulations based on density-functional theory (DFT) are routinely performed for surfaces of crystals and crystalline nanoparticles.[1−3] The issue is more complicated, however, for amorphous matter: the absence of translational symmetry means that there are no well-defined cleavage planes nor ordered unit cells. Atomistic modeling of amorphous-material surfaces therefore requires large structural models, long molecular-dynamics (MD) simulations, and accurate energies and interatomic forces, all at the same time. At present, this is impossible to achieve for a quantum-mechanical method, such as DFT.

Diamond-like or "tetrahedral" amorphous carbon (*ta*-C), a dense form of carbon with a large concentration of fourfold-coordinated atoms and a density around 3 g cm^{-3}, is an important example of amorphous functional materials.[4−6] Due to its attractive mechanical properties, *ta*-C is widely used in coatings (for example, to protect machinery from mechanical wear) and tribological applications.[6−8] It has also been suggested and used as an electrode material for detecting biomolecules.[9−11] All these applications are directly enabled by atomic-scale adsorption and reaction mechanisms at the amorphous surfaces and by the interfaces they form with chemical environments. Accurate modeling of *ta*-C surfaces and understanding their chemical reactivity is therefore crucial for the design of *ta*-C materials.

Over the years, several key papers reported simulations of *ta*-C surfaces, but these have mainly been based on classical (empirically fitted) force fields. For example, the widely used interatomic potential by Tersoff was initially applied to amorphous carbon surfaces.[12] MD simulations of *ta*-C film growth were reported, directly mimicking the impact of atoms on a substrate.[13−17] Still, these classical potentials face inherent challenges: surface energies of diamond and *ta*-C can be difficult to describe,[18,19] and various potentials describe the high-temperature "graphitization" of *ta*-C in qualitatively different ways.[20] In principle, DFT could provide the required accuracy for these tasks, but it quickly becomes prohibitively expensive for large simulation cells. Consequently, previous DFT studies of *ta*-C surfaces have been limited to relatively small model system sizes.[21−25]

Machine-learning (ML)-based interatomic potentials are emerging tools for materials simulations, promising to alleviate some of these problems. By "learning" (interpolating) high-dimensional potential-energy surfaces, ML-based potentials enable atomistic simulations at close-to-DFT accuracy (to within a few kJ/mol) but with a computational cost that is several orders of magnitude lower.[26−31] Recent studies suggest that ML-based potentials are becoming viable tools for materials chemistry and physics.[32−41] We recently used such a potential for large-scale deposition simulations of *ta*-C films, describing the impact of thousands of individual atoms, one at a time, and achieving excellent agreement with experimental observables (including the count of fourfold-coordinated"sp^3" atoms and mechanical properties).[38]

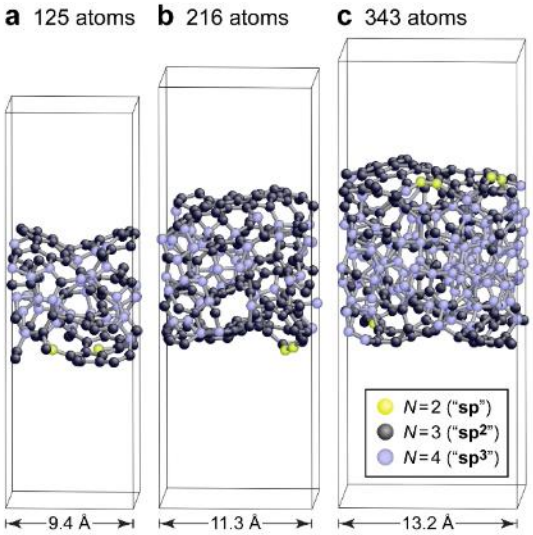

Figure 1. Exemplary structure models for *ta*-C surfaces, obtained in GAP-driven MD simulations. Atoms are colored according to their number of neighbors, *N* (counted using a 1.85 Å cutoff, which corresponds to the first minimum of the radial distribution function): two (*yellow*), three (*dark gray*), and four (*blue*). Thin lines denote the boundaries of simulation cells: the structures extend in two dimensions and are terminated by vacuum in the third, exposing *ta*-C surfaces at top and bottom.

While most simulations so far have dealt with surfaces of the pure amorphous element, a further key role is played by chemical functionalization: that is, by the presence of hydrogen, oxygen, and other atomic species and functional groups. Such an extension of compositional space is challenging for any ML-based potential because it requires large amounts of additional reference data. In contrast, arbitrary chemical compositions can trivially be entered in first-principles DFT as long as the system size is sufficiently small. The methods therefore do not exclude but complement each other: once an accurate slab model of the pure surface has been obtained in fast ML-driven simulations, it is amenable to further DFT studies. We have shown the usefulness of such combined ML/DFT approaches very recently for Na intercalation in bulk disordered carbon structures.[42]

In this work, we propose combining ML and established simulation methods to study structures and the reactivity of amorphous material surfaces, as applied here to *ta*-C. We create and validate a library of optimized "slab" models (Figure 1) that enable a range of atomistic studies from first principles. We study the generation of progressively hydrogenated *ta*-C:H slabs using a heuristic Monte Carlo (MC) scheme and interactions obtained from Density-Functional Tight-Binding (DFTB),[43,44] and we perform DFT-MD studies of the atomic-scale oxidation mechanisms: in other words, we move beyond pure *ta*-C both to hydrogen and oxygen surface species. This opens the door for the realistic modeling and understanding of amorphous material surfaces, including their chemical functionalization, with unprecedented accuracy.

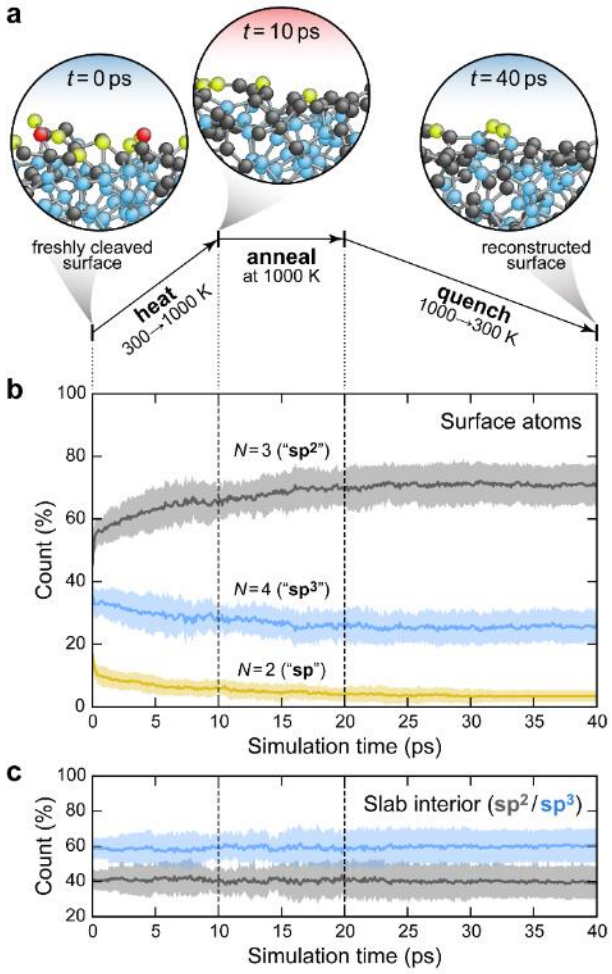

Figure 2. Generating optimized *ta*-C surface models. (a) Sketch of the temperature profile used to anneal structures after cleaving from the bulk, as in ref 19, with exemplary structural snapshots shown. Atoms are colored according to their number of neighbors: one (*red*), two (*yellow*), three (*dark gray*), and four (*blue*). (b) Statistics of atomic coordination at the surfaces (topmost 3 Å; see Supporting Information), measuring the degree of graphitization during simulations. Data have been averaged over the respective most stable slabs from each of the bulk 216-atom structures (10 in total); shading indicates standard deviations. (c) Same for the concentration of sp^2 and sp^3 atoms within the reconstructed slabs (central 3 Å), evidencing that the central regions remain bulk-like and do not reconstruct on annealing. The results for the slab interior are consistent with bulk melt–quench simulations using GAP.[19] The experimental sp^3 content of *ta*-C is even larger (> 80%), which can be reproduced by explicit deposition simulations using GAP-MD,[38] but these require much larger system sizes (> 10,000 atoms); for a discussion, see ref 38.

RESULTS AND DISCUSSION

Surface Reconstructions. Our simulation protocol and its main outcomes are sketched in Figure 2. We begin by "cleaving" slabs from the bulk: as is common practice in crystalline systems,[1–3] we take *ta*-C bulk structures, here containing 216 atoms (from ref 19), and insert an artificial vacuum region in one dimension (*cf.* Figure 1b). We use 10 uncorrelated cells and cleave each at five equidistant planes to enhance the sampling. Using MD simulations driven by our ML-based Gaussian approximation potential (GAP),[19]

we then heat and anneal all 50 slabs at 1,000 K. These simulations induce surface reconstructions and quickly "heal" the dangling-bond defects (red atoms in Figure 2a).[19]

The most common (and simplistic) way to quantify the local structure in *ta*-C is to count three- and four-fold coordinated atoms ("sp^2/sp^3"). We use this to monitor structural transformations at the surfaces (Figure 2b): during heating and annealing, the connectivity of near-surface atoms changes visibly, and graphitic "sp^2" atoms are created (*gray*) while chain-like "sp" atoms (*yellow*) are gradually removed. This is consistent with surface "graphitization" as described earlier.[13,14,19] We also verify that the *interiors* of our slabs remain bulk-like during annealing, as analyzed using both the sp^2/sp^3 count (Figure 2c) and the mass density (Figure S2).

System Size and Structural Metrics. While GAP-MD can easily generate *ta*-C structures with thousands of atoms in the unit cell (since GAP is a linearly scaling potential),[19,38] we focus here on smaller systems on purpose: we want to analyze them using DFTB or DFT subsequently, as said above. Hence, what is the smallest system size that can still give a physically meaningful description of *ta*-C surfaces? To assess this question, we generated more slab models with different numbers of atoms: first, quenching bulk *ta*-C from the melt as in ref 19, then, cleaving and annealing surface structures (*cf.* Figure 2). Simulations with 512 atoms per cell serve as the benchmark, as this is currently the largest system size routinely accessible to subsequent DFT modeling. Again, we look at two key structural indicators: the sp^3 count in the slab centers (Figure 3a) and the sp^2 count at the surfaces (Figure 3b), defined as above. For both measures, with 125 atoms already, the mean values largely converge to within the standard deviation of the 512-atom benchmark (shaded). In contrast, 64-atom systems do not give reliable results, and in one case the entire slab graphitized, reflected in an unphysically low sp^3 count.

Characterizing carbon structures by counting nearest neighbors is intuitive but has intrinsic shortcomings. For example, square-planar coordinated carbon is a highly energetic and elusive species,[45] but having four neighbors it would be assigned as "sp^3" using the above scheme. Improved structural metrics based on bond angles[46,47] have been applied to carbon.[48,49] We here take a more general approach originating in the ML field: namely, a numerical similarity measure, or *kernel*, to compare atomic environments. We use the "Smooth Overlap of Atomic Positions" (SOAP),[50] as detailed in the Methods section. SOAP has proven useful both for fitting GAPs[19,51] and for analyzing chemical structures.[52,53] Here, we show that the approach is particularly beneficial for amorphous surfaces, as it readily identifies special surface sites and enables a quantitative comparison of bulk and surface structures.

SOAP measures the similarity for a given pair of atoms, normalized to a range between zero (fully dissimilar) and unity (identical within a given cutoff). We start by calculating this quantity for each atom in a *ta*-C slab as compared to an atom in bulk diamond (Figure 3c). As expected, the slab interior is structurally similar to diamond, albeit with pronounced scatter (due to the "sp^2/sp^3" coexistence, reflected by colors ranging from light to dark green). The surface regions are much less similar to diamond, in line with the graphitization discussed before, and the lower coordination numbers overall (Figure 3a–b). In turn, repeating the SOAP analysis but now comparing to graphene shows clear similarity at the surfaces but not in the bulk (Figure 3d). The approach also readily identifies a surface "sp" motif that is dissimilar both to diamond and graphene (–C≡C–; arrows).

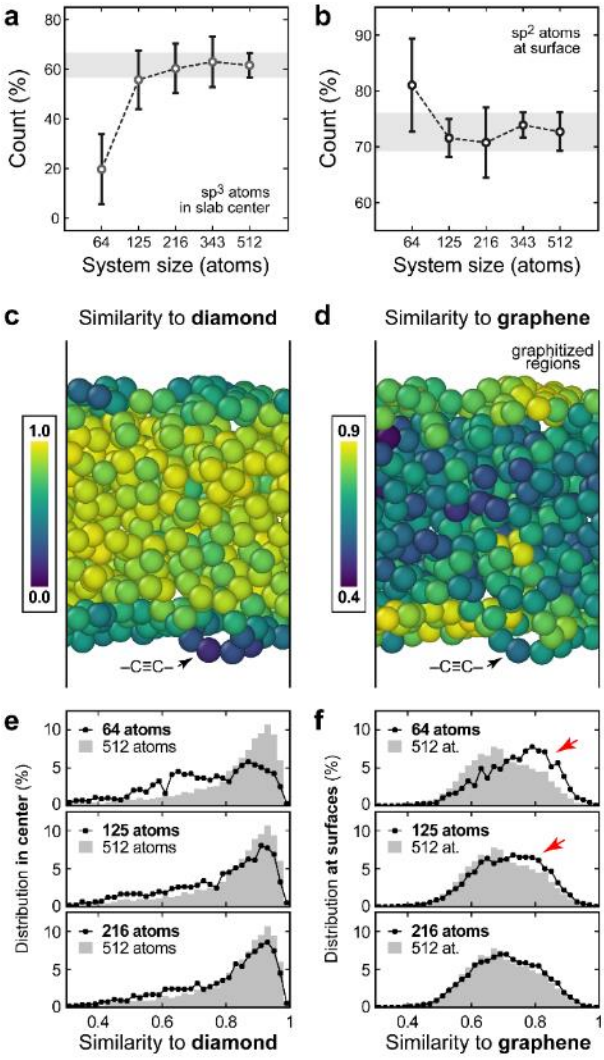

Figure 3. System-size dependence of *ta*-C slab models and a quantitative measure for their local structure. (a) Count of sp^3 atoms in the center of the slabs, as in Figure 2c, but now as a function of system size. (b) Same for the sp^2 count at the surfaces. Error bars give standard deviations. (c) Structural analysis using SOAP. A *ta*-C surface slab containing 512 atoms is shown; atoms are colored according to their SOAP similarity to diamond, the prototype for "sp^3" carbon. Yellow atoms are highly similar to diamond (in the center of the slab); blue atoms are dissimilar (mainly near the surface). (d) Same for a comparison to graphene as a prototype for "sp^2" carbon. (e) Fingerprints of SOAP similarity to diamond for different system sizes, obtained by collecting data over atoms in the center as above. Values lower than 0.3 are very rare and the horizontal axis is thus truncated to ease comparison. (f) Same but for SOAP similarity to graphene, measured in the surface regions as above. Arrows highlight deviations from the reference (512 atoms).

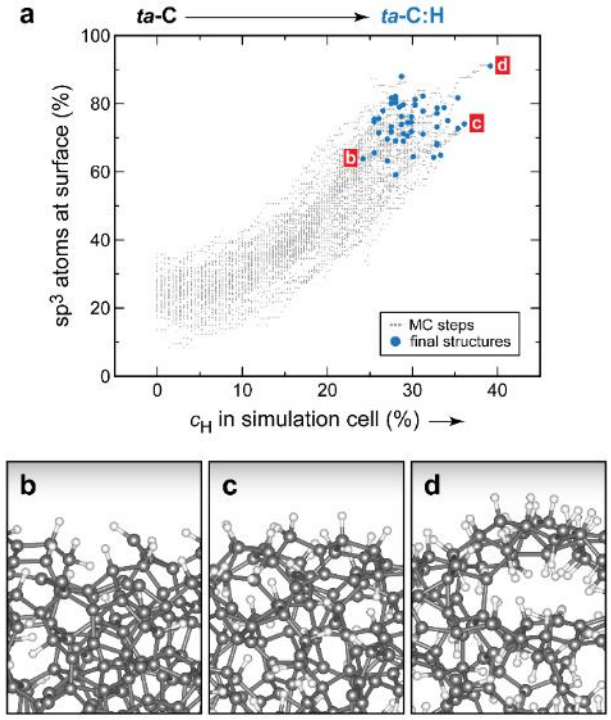

Figure 4. Hydrogenation of *ta*-C surfaces, to give "*ta*-C:H", studied using Monte Carlo (MC) simulations as detailed in the Methods section. (a) Count of carbon atoms in fourfold bonded "sp³" configurations on the surface (topmost 3 Å) for simulation cells with progressively increasing hydrogen content. Each gray dot corresponds to one accepted step out of 50 MC trajectories, and each blue dot corresponds to one of the end points. (b–d) Structural snapshots from selected final structures with different c_H, as marked above in red. Panel (d) shows an extreme case where subsurface voids developed but the slab still stayed intact.

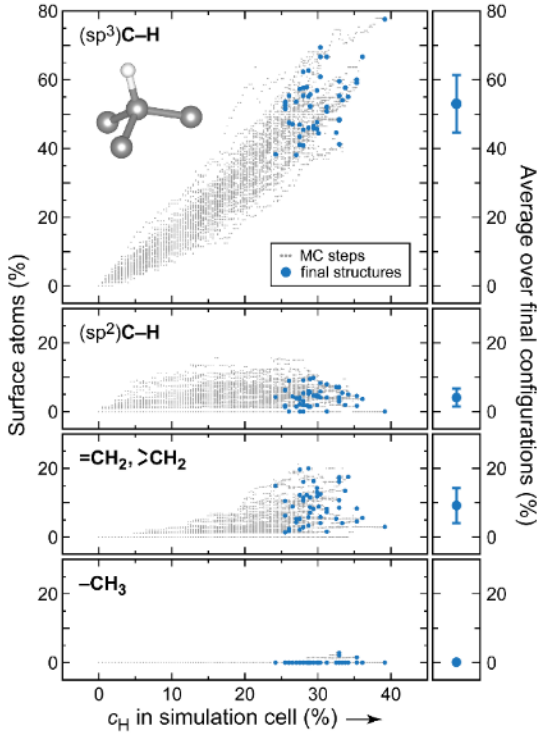

Figure 5. Evolution of different hydrogen-containing environments during the GCMC simulations (*main panels*), and the averaged abundance of these motifs in the final hydrogenated structures (*right-hand side*) on the surface (topmost 3 Å). Error bars, where reasonable, indicate standard deviations over 50 simulation cells.

With this algorithm, we can now obtain numerical fingerprints to better characterize our structural models. We collect histograms, similar in spirit to ref 54, but now using SOAP. Again, we restrict our analysis of "diamond-like-ness" (Figure 3e) to the central regions of the slabs, and that of surface graphitization (Figure 3f) to the outermost 3 Å. The reference distributions (512 atoms; gray shading) peak at ≈ 0.9 for diamond similarity in the bulk, and at ≈ 0.7 for graphene similarity at the surface. We observe a few sites with formally high similarity (≈0.9) to *both* crystalline structures; we stress, however, that the numerical scales are somewhat arbitrary, and the focus should be on the *relative* similarity of different sites in a given structure. Returning to the question of system-size dependence, clear deviations from the reference are seen for the surfaces in the 64- and 125-atom systems (arrows in Figure 3f), suggesting that both are structurally different—a dissimilarity that, for the 125-atom systems, could not have been spotted by counting "sp²/sp³" environments alone (Figure 3a–b). By contrast, at 216 atoms, the SOAP fingerprints, both in the center and at the surface, agree very well with the respective benchmark. We therefore use 216-atom systems for subsequent DFTB/DFT studies of *ta*-C surface chemistry.

Beyond the present work, the smaller structures (125 atoms) could still be useful for higher-level computations of electronic properties, and the larger ones (343–512 atoms) for studying mechanical properties. All surface models generated in this work are provided as Supporting Information.

Surface Hydrogenation. Incorporating hydrogen in amorphous carbon phases leads to diverse materials with interesting properties.[6,55] The most relevant to the present work is so-called hydrogenated *ta*-C, or "*ta*-C:H", which was experimentally shown to exhibit large hardness and sp³ concentration (≈ 75% of carbon atoms) at a hydrogen content of up to ≈ 30%.[56] To study hydrogenation of *ta*-C surfaces in a representative ensemble of structures, we decided to perform Markov-chain Monte Carlo (MC) simulations of hydrogenation, exploring diverse local environments and hydrogen contents. As the underlying *ta*-C is already reliably described by the GAP,[19,38] and the modeling of *local* C–H interactions is a problem of limited chemical complexity, we resort to the DFTB method[43,44] which is much less computationally demanding than DFT, thus allowing us to carry out 50 MC trajectories in parallel. Below, we will also address a problem of larger chemical complexity (namely, oxidation), for which we will use first-principles DFT-MD instead.

Starting from 50 unique, optimized 216-atom slabs introduced above, we gradually added hydrogen atoms to each and accepted or rejected each individual step based on an energy criterion (Methods section). Figure 4a characterizes the outcome of these MC trajectories as a function of the hydrogen concentration, $c_H = N_H / (N_C + N_H)$: each data point refers to an accepted step in one of the cells, and the surface concentration of fourfold bonded "sp³" atoms is plotted on the *y*-axis (again, considering as surface region the topmost 3 Å). The latter increases with c_H, in line with the simplified chemical notion of double-bonded "sp²" environments being saturated by hydrogenation, forming "sp³" sites. In the

final structures, 74±6% of surface carbon atoms are fourfold bonded (blue points in Figure 4a); there is, of course, some scatter due to the necessarily finite system sizes. Snapshots from three simulation cells are shown in Figure 4b–d, visualizing interesting limiting cases in terms of c_H.

The local chemical environments can be further analyzed by collecting statistics about the numbers of C/H neighbor atoms (Figure 5). The most prominent motif we find is an "sp³" carbon atom bonded to three C and one H neighbors, with an abundance of 53±8% in the final structures. Another important role, although only at larger hydrogen content, is played by doubly hydrogenated CH_2 units (9±5%); several of these are seen in Figure 4d, partly explaining the very high surface sp³ content (> 90%) observed for this particular structure. In contrast, methyl groups ($-CH_3$) are not formed in significant amounts in our MC simulations.

Surface Oxidation. We finally move on to the modeling of oxygen at *ta*-C surfaces, which represents the second key type of functionalization—and arguably the more complicated one. Somewhat reminiscent of organic chemistry, "ether", "keto", and other motifs have been discussed in an early DFT study of ideal[57] and worn[58,59] diamond surfaces in contact with oxygen. One would expect similar groups to be present at the amorphous surfaces, but here the structural variety is much larger. While the existence of various oxygen-based surface species has been observed in X-ray absorption spectroscopy (XAS)[60] and X-ray photoelectron spectroscopy (XPS),[61,62] no structural information is known *a priori*. We therefore decided, in this case, to perform DFT-MD simulations to gain insight from first principles.

We randomly selected six slabs, randomly placed either four or eight oxygen atoms on either side and heated the systems from 300 K to 1,800 K (in 300 K increments). The holding time at each increment was 5 ps (≡ 5,000 steps). Exemplary structural snapshots from these DFT-MD trajectories are shown in Figure 6a–c. For one sample, we heated the slab further to 3,000 K; at these temperatures, the structures disintegrate by releasing CO (*cf.* Figure 6c). We did not observe formation of CO_2, which we ascribe to the relatively low concentration of oxygen on the surfaces. We also performed additional tests with further annealing (10 ps at each temperature increment) for one of the systems.

Overall, our exploratory DFT-MD simulations suggest the presence of three main surface species (Figure 6d); recall that the initial adsorption sites are sampled at random, without chemical preconceptions. We quantify the abundance of sites, based on the atomic connectivity, and report the results in Table 1: these data still comprise a limited sample size (72 independent oxygen environments), but nonetheless provide clear chemical insight. Epoxy groups **1** are abundant early on in our simulations, but during annealing their count decreases progressively, presumably due to the high strain in this three-membered ring structure, and consistent with the high reactivity of epoxy groups in organic molecules. Ether groups **2** are less abundant in the 300 K simulations, but they persist throughout annealing. Keto groups **3** are the most common structural fragment found in our simulations, qualitatively consistent with the XAS result from ref 60, and become even more important with increasing annealing temperature.

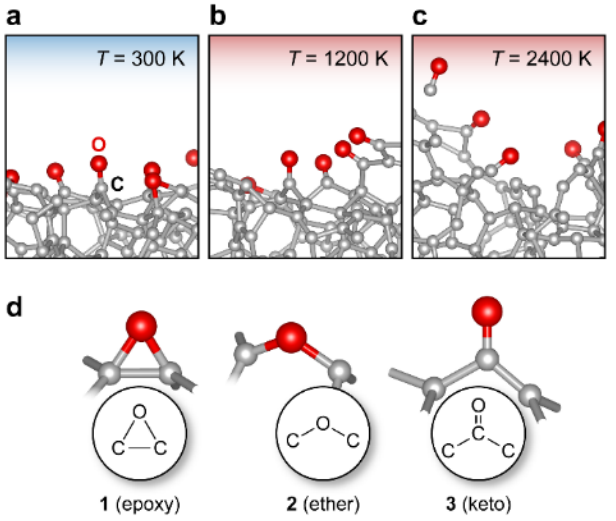

Figure 6. Surface oxidation of *ta*-C. (a–c) Exemplary structural snapshots during progressive DFT-MD annealing of the same *ta*-C surface slab with chemisorbed oxygen atoms. At low to moderate temperatures, mostly epoxy and keto groups are found, and no significant structural changes occur (panel a). At 1,200 K, the epoxy rings have been broken and surface reconstruction takes place (panel b). At much higher temperatures, CO molecules detach from the surface (panel c). (d) Characteristic oxygen-containing surface motifs, as observed in our simulations: example snapshots from a 300 K trajectory are shown.

Table 1. Count of Functional Groups at *ta*-C Surfaces as Observed in DFT-MD Simulations

T (K)	**1** (epoxy)	**2** (ether)	**3** (keto)	Other
300	27	13	28	4
600	19	13	34	6
900	16	17	35	4
1200	12	18	35	7
1500	7	18	39	8
1800	2	18	41	11

We note that MD simulations using empirical potentials, which are less accurate than DFT but applicable to larger systems, have been reported for chemically reduced graphene oxide (rGO).[63–65] We envision that further insight into both materials classes will be possible by exploiting their close chemical similarity (quantitatively supported by the graphene-likeness at *ta*-C surfaces; Figure 3d). We also note that temperature-induced decrease in epoxy groups **1** and increase in keto groups **3** was observed in MD simulations for rGO,[63] qualitatively in line with our results in Table 1.

With the onset of decarbonylation of the slab, the two most relevant structural motifs observed in our MD simulations are keto groups **3** and allylic, chain-like –C=C=O motifs; on heating even further, CO molecules detach from the slab (Figure 6c). The latter agrees with previous DFT studies of formation and oxidation of linear carbon chains;[58,59] these studies, however, were conducted on small model systems (such as linear molecules), not on slab cells. Further ongoing work will deal with the following questions: (i) how can one automatically identify and catalogue the most representative surface sites (using the SOAP kernel in the spirit

of ref 52 and Figure 3c–d);[66] and (ii) what are the spectroscopic fingerprints of these particular surface sites?

CONCLUSIONS

Machine learning combined with DFT modeling can give new insights into the surface chemistry of amorphous materials. We exemplified this here for tetrahedral amorphous carbon (*ta*-C), one of the prototypical amorphous materials whose surfaces are of key technological importance. We generated and characterized a representative set of slab models that can enable future DFT studies at various system sizes, and we assembled large databases of functionalized structures that may serve for fitting multicomponent potential models in the future. Direct comparison to experiments, including spectroscopic probes such as XAS/XPS, will ultimately make it possible to quantify, understand, and control the surface chemistry of *ta*-C. Looking beyond carbon, with the recent development of high-quality machine-learning potentials for other (bulk) amorphous solids,[36,67] we expect that simulation strategies as introduced here will be useful for other amorphous material surfaces in the near future.

METHODS

GAP-MD simulations in the NVT ensemble were used to generate *ta*-C surface slabs, using similar protocols as in our initial work.[19] Simulations were performed using a Langevin thermostat as implemented in QUIP/quippy, which is freely available for non-commercial research at https://github.com/libAtoms/QUIP. The GAP approach (ref 27) has been validated earlier for a range of relevant applications, including the description of metal surfaces,[51] structure-searching for carbon allotropes,[49] and lithium diffusion in graphitic systems.[37]

Borrowing ideas from Grand-canonical Monte Carlo, our heuristic Markov-chain MC simulations (Figure 4–5) progressed by randomly placing H atoms near undercoordinated (less than four-fold bonded) C atoms or removing H atoms (if present) and then relaxing the system to its nearest local minimum.[68] In each step, the choice to add or remove H was made with 50% probability, and the specific insertion or removal site was then chosen with equal probability among the possible sites. Trial moves were accepted with a Metropolis criterion[69] based on the energy difference between initial and final states. Specifically, given the (total) energy E_n for a system with n chemisorbed H atoms and m possible adsorption sites, we accept the addition of an H atom with probability

$$P_{+1} = \min\left\{1, \frac{n+1}{m}\exp\left[-\frac{E_{n+1}-(E_n+\mu)}{k_BT}\right]\right\}$$

and accept the removal of an H atom with probability

$$P_{-1} = \min\left\{1, \frac{m+1}{n}\exp\left[-\frac{E_{n-1}-(E_n-\mu)}{k_BT}\right]\right\},$$

where μ is half the binding energy of molecular H_2 in our DFTB model, $\mu = -2.6\text{eV}$. Therefore, $E_{\text{react}} = -E_{n+1} + (E_n + \mu)$ is the reaction energy of the corresponding hydrogenation reaction. The factor in front of the exponential accounts for the fact that the number of adsorbed H and the number of potential adsorption sites change during the course of the MC simulation. All simulations were run at $T = 300\text{K}$, but since $k_BT \ll \langle E_{\text{react}}\rangle$, the results obtained here are largely independent of temperature for typical ambient temperatures.

We note that this scheme is heuristic because it does not fulfill detailed balance, since addition of hydrogen can trigger fold catastrophes that relax to a new minimum energy configuration. Initial parametric tests using the Reactive Empirical Bond Order Potential (REBO2)[70] showed that the resulting structures were largely independent of the temperature chosen in the MC scheme; this is because the hydrogen adsorption energy is far larger than k_BT for T smaller than the melting point of diamond. All results presented here were obtained at 300 K. Our computations do not include vibrational contributions to the free energy, and do not include other gaseous species (such as H_2O in a humid atmosphere); both constitute interesting further directions but will increase the computational cost very significantly. Final simulations employed the DFTB method[44] with the mio-1-1 parameter set,[44] available at http://dftb.org, without self-consistent charges as implemented in the ATOMISTICA code, which is freely available at https://github.com/Atomistica/atomistica. Each of the 50 trajectories ran for 1000 steps of adding/removing hydrogen from the structures.

DFT-MD simulations of *ta*-C surface oxidation (Figure 6) were performed using the QUICKSTEP algorithm as implemented in CP2K.[71–73] We used double-ζ quality basis sets, Goedecker–Teter–Hutter pseudopotentials,[74] the PBE functional for exchange and correlation,[75] without spin polarization, and a cut-off energy of 250 Ry. Reciprocal space was sampled at Γ. Temperature was controlled using a stochastic Langevin thermostat.[76] The timestep, both for GAP-MD and DFT-MD simulations, was 1 fs.

Structural similarity to diamond and graphene (Figure 3c–f) was quantified using the average Smooth Overlap of Atomic Positions ("average SOAP") kernel.[50,52] SOAP is a numerical similarity measure based on the neighborhood density of atoms expanded into spherical harmonics: we use convergence parameters of $n_{\text{max}} = l_{\text{max}} = 10$, a radial cutoff of 3.5 Å, a selectivity parameter of $\zeta = 4$, and a smoothness parameter of $\sigma_{\text{at}} = 0.5$ Å (the latter setting controls how "fuzzy" the neighbor densities are, and is the same as used for fitting the GAP model). The histograms given in Figures 3e–f represent averages over 50 different slab structures for each given system size. For visualization, we used AtomEye,[77] OVITO,[78] and VESTA.[79]

ASSOCIATED CONTENT

Supporting Information. Archives containing structural data, in XYZ format, for the optimized pristine (*cf.* Figures 1–3), hydrogenated (*cf.* Figures 4–5), and oxidized (*cf.* Figure 6) *ta*-C slab models generated in this work (ZIP)
Supplementary discussion, including radial distribution function (RDF) analysis, mass density profiles, and additional discussion regarding Monte Carlo simulations (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: vld24@cam.ac.uk.

ACKNOWLEDGEMENTS

V.L.D. gratefully acknowledges a Feodor Lynen fellowship from the Alexander von Humboldt Foundation, a Leverhulme Early Career Fellowship, and support from the Isaac Newton Trust. T.L., A.A., and M.A.C. acknowledge funding from the Academy of Finland through grants no. 285526 and 310574. R.J. and L.P. acknowledge funding by the Deutsche Forschungsgemeinschaft (grant PA 2023/2). This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730897 (HPC-Europa3). The authors acknowledge CSC – IT Center for Science, Finland, the Jülich Supercomputing Center (project "hfr13"), the state of Baden-Württemberg (project bwHPC) and the Deutsche For-

schungsgemeinschaft (grant INST 39/963-1 FUGG) for computational resources. Access to the ARCHER UK National Supercomputing Service was provided via EPSRC Grant EP/P022596/1. *Data access statement:* Data supporting this publication are available as electronic Supporting Information.

REFERENCES

(1) Groß, A. *Theoretical Surface Science: A Microscopic Perspective*; Springer: Berlin, Heidelberg, New York, 2009.

(2) Barnard, A. S. Modelling of Nanoparticles: Approaches to Morphology and Evolution. *Rep. Prog. Phys.* **2010**, *73*, 86502.

(3) Deringer, V. L.; Dronskowski, R. From Atomistic Surface Chemistry to Nanocrystals of Functional Chalcogenides. *Angew. Chem. Int. Ed.* **2015**, *54*, 15334–15340.

(4) McKenzie, D.; Muller, D.; Pailthorpe, B. Compressive-Stress-Induced Formation of Thin-Film Tetrahedral Amorphous Carbon. *Phys. Rev. Lett.* **1991**, *67*, 773–776.

(5) McKenzie, D. R. Tetrahedral Bonding in Amorphous Carbon. *Reports Prog. Phys.* **1999**, *59*, 1611–1664.

(6) Robertson, J. Diamond-like Amorphous Carbon. *Mater. Sci. Eng. R Rep.* **2002**, *37*, 129–281.

(7) Bewilogua, K.; Hofmann, D. History of Diamond-like Carbon Films — From First Experiments to Worldwide Applications. *Surf. Coatings Technol.* **2014**, *242*, 214–225.

(8) Music, D.; Geyer, R. W.; Schneider, J. M. Recent Progress and New Directions in Density Functional Theory Based Design of Hard Coatings. *Surf. Coatings Technol.* **2016**, *286*, 178–190.

(9) Zeng, A.; Neto, V. F.; Gracio, J. J.; Fan, Q. H. Diamond-like Carbon (DLC) Films as Electrochemical Electrodes. *Diam. Relat. Mater.* **2014**, *43*, 12–22.

(10) Palomäki, T.; Chumillas, S.; Sainio, S.; Protopopova, V.; Kauppila, M.; Koskinen, J.; Climent, V.; Feliu, J. M.; Laurila, T. Electrochemical Reactions of Catechol, Methylcatechol and Dopamine at Tetrahedral Amorphous Carbon (ta-C) Thin Film Electrodes. *Diam. Relat. Mater.* **2015**, *59*, 30–39.

(11) Laurila, T.; Sainio, S.; Caro, M. A. Hybrid Carbon Based Nanomaterials for Electrochemical Detection of Biomolecules. *Prog. Mater. Sci.* **2017**, *88*, 499–594.

(12) Tersoff, J. Empirical Interatomic Potential for Carbon, with Application to Amorphous Carbon. *Phys. Rev. Lett.* **1988**, *61*, 2879–2882.

(13) Marks, N. A.; Cover, M. F.; Kocer, C. Simulating Temperature Effects in the Growth of Tetrahedral Amorphous Carbon: The Importance of Infrequent Events. *Appl. Phys. Lett.* **2006**, *89*, 131924.

(14) Marks, N.; Cover, M.; Kocer, C. The Importance of Rare Events in Thin Film Deposition: A Molecular Dynamics Study of Tetrahedral Amorphous Carbon. *Mol. Simul.* **2006**, *32*, 1271–1277.

(15) Jäger, H. U.; Albe, K. Molecular-Dynamics Simulations of Steady-State Growth of Ion-Deposited Tetrahedral Amorphous Carbon Films. *J. Appl. Phys.* **2000**, *88*, 1129–1135.

(16) Jäger, H. U.; Belov, A. Y. Ta-C Deposition Simulations: Film Properties and Time-Resolved Dynamics of Film Formation. *Phys. Rev. B* **2003**, *68*, 24201.

(17) Belov, A. Y.; Jäger, H. U. Calculation of Intrinsic Stresses in Amorphous Carbon Films Grown by Molecular Dynamics Simulation: From Atomic to Macroscopic Scale. *Comput. Mater. Sci.* **2002**, *24*, 154–158.

(18) Pastewka, L.; Klemenz, A.; Gumbsch, P.; Moseler, M. Screened Empirical Bond-Order Potentials for Si-C. *Phys. Rev. B* **2013**, *87*, 205410.

(19) Deringer, V. L.; Csányi, G. Machine Learning Based Interatomic Potential for Amorphous Carbon. *Phys. Rev. B* **2017**, *95*, 094203.

(20) de Tomas, C.; Suarez-Martinez, I.; Marks, N. A. Graphitization of Amorphous Carbons: A Comparative Study of Interatomic Potentials. *Carbon* **2016**, *109*, 681–693.

(21) Dong, J.; Drabold, D. A. Ring Formation and the Structural and Electronic Properties of Tetrahedral Amorphous Carbon Surfaces. *Phys. Rev. B* **1998**, *57*, 15591–15598.

(22) Kaukonen, M.; Nieminen, R. M.; Pöykkö, S.; Seitsonen, A. P. Nitrogen Doping of Amorphous Carbon Surfaces. *Phys. Rev. Lett.* **1999**, *83*, 5346–5349.

(23) Bauschlicher, Jr., C. W.; Lawson, J. W. Amorphous Carbon and Its Surfaces. *Chem. Phys.* **2010**, *374*, 77–82.

(24) Caro, M. A.; Zoubkoff, R.; Lopez-Acevedo, O.; Laurila, T. Atomic and Electronic Structure of Tetrahedral Amorphous Carbon Surfaces from Density Functional Theory: Properties and Simulation Strategies. *Carbon* **2014**, *77*, 1168–1182.

(25) Aarva, A.; Laurila, T.; Caro, M. A. Doping as a Means to Probe the Potential Dependence of Dopamine Adsorption on Carbon-Based Surfaces: A First-Principles Study. *J. Chem. Phys.* **2017**, *146*, 234704.

(26) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(27) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

(28) Li, Z.; Kermode, J. R.; De Vita, A. Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces. *Phys. Rev. Lett.* **2015**, *114*, 096405.

(29) Shapeev, A. Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Multiscale Model. Simul.* **2016**, *14*, 1153–1173.

(30) Huan, T. D.; Batra, R.; Chapman, J.; Krishnan, S.; Chen, L.; Ramprasad, R. A Universal Strategy for the Creation of Machine Learning-Based Atomistic Force Fields. *npj Comput. Mater.* **2017**, *3*, 37.

(31) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192–3203.

(32) Khaliullin, R. Z.; Eshet, H.; Kühne, T. D.; Behler, J.; Parrinello, M. Nucleation Mechanism for the Direct Graphite-to-Diamond Phase Transition. *Nat. Mater.* **2011**, *10*, 693–697.

(33) Elias, J. S.; Artrith, N.; Bugnet, M.; Giordano, L.; Botton, G. A.; Kolpak, A. M.; Shao-Horn, Y. Elucidating the Nature of the Active Phase in Copper/Ceria Catalysts for CO Oxidation. *ACS Catal.* **2016**, *6*, 1675–1679.

(34) Morawietz, T.; Singraber, A.; Dellago, C.; Behler, J. How van Der Waals Interactions Determine the Unique Properties of Water. *Proc. Natl. Acad. Sci., U. S. A.* **2016**, *113*, 8368–8373.

(35) Gabardi, S.; Baldi, E.; Bosoni, E.; Campi, D.; Caravati, S.; Sosso, G. C.; Behler, J.; Bernasconi, M. Atomistic Simulations of the Crystallization and Aging of GeTe Nanowires. *J. Phys. Chem. C* **2017**, *121*, 23827–23838.

(36) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine Learning Unifies the Modeling of Materials and Molecules. *Sci. Adv.* **2017**, *3*, e1701816.

(37) Fujikake, S.; Deringer, V. L.; Lee, T. H.; Krynski, M.; Elliott, S. R.; Csányi, G. Gaussian Approximation Potential Modeling of Lithium Intercalation in Carbon Nanostructures. *J. Chem. Phys.* **2018**, *148*, 241714.

(38) Caro, M. A.; Deringer, V. L.; Koskinen, J.; Laurila, T.; Csányi, G. Growth Mechanism and Origin of High sp^3 Content in Tetrahedral Amorphous Carbon. *Phys. Rev. Lett.* **2018**, *120*, 166101.

(39) Artrith, N.; Urban, A.; Ceder, G. Constructing First-Principles Phase Diagrams of Amorphous Li_xSi Using Machine-Learning-Assisted Sampling with an Evolutionary Algorithm. *J. Chem. Phys.* **2018**, *148*, 241711.

(40) Onat, B.; Cubuk, E. D.; Malone, B. D.; Kaxiras, E. Implanted Neural Network Potentials: Application to Li-Si Alloys. *Phys. Rev. B* **2018**, *97*, 094106.

(41) Deringer, V. L.; Bernstein, N.; Bartók, A. P.; Kerber, R. N.; Cliffe, M. J.; Marbella, L. E.; Grey, C. P.; Elliott, S. R.; Csányi, G. Realistic Atomistic Structure of Amorphous Silicon from Machine-Learning-Driven Molecular Dynamics. *J. Phys. Chem. Lett.* **2018**,

9, 2879–2885.

(42) Deringer, V. L.; Merlet, C.; Hu, Y.; Lee, T. H.; Kattirtzi, J. A.; Pecher, O.; Csányi, G.; Elliott, S. R.; Grey, C. P. Towards an Atomistic Understanding of Disordered Carbon Electrode Materials. *Chem. Commun.* **2018**, *54*, 5988–5991.

(43) Frauenheim, T.; Weich, F.; Köhler, T.; Uhlmann, S.; Porezag, D.; Seifert, G. Density-Functional-Based Construction of Transferable Nonorthogonal Tight-Binding Potentials for Si and SiH. *Phys. Rev. B* **1995**, *52*, 11492–11501.

(44) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-Consistent-Charge Density-Functional Tight-Binding Method for Simulations of Complex Materials Properties. *Phys. Rev. B* **1998**, *58*, 7260–7268.

(45) Hoffmann, R.; Alder, R. W.; Wilcox, C. F. Planar Tetracoordinate Carbon. *J. Am. Chem. Soc.* **1970**, *92*, 4992–4993.

(46) Chau, P.-L.; Hardwick, A. J. A New Order Parameter for Tetrahedral Configurations. *Mol. Phys.* **1998**, *93*, 511–518.

(47) Yang, L.; Powell, D. R.; Houser, R. P. Structural Variation in Copper(I) Complexes with Pyridylmethylamide Ligands: Structural Analysis with a New Four-Coordinate Geometry Index, τ_4. *Dalt. Trans.* **2007**, No. 9, 955–964.

(48) Öhrström, L.; O'Keeffe, M. Network Topology Approach to New Allotropes of the Group 14 Elements. *Z. Krist.* **2013**, *228*, 343–346.

(49) Deringer, V. L.; Csányi, G.; Proserpio, D. M. Extracting Crystal Chemistry from Amorphous Carbon Structures. *ChemPhysChem* **2017**, *18*, 873–877.

(50) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B* **2013**, *87*, 184115.

(51) Szlachta, W. J.; Bartók, A. P.; Csányi, G. Accuracy and Transferability of Gaussian Approximation Potential Models for Tungsten. *Phys. Rev. B* **2014**, *90*, 104108.

(52) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing Molecules and Solids across Structural and Alchemical Space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.

(53) Mavračić, J.; Mocanu, F. C.; Deringer, V. L.; Csányi, G.; Elliott, S. R. Similarity between Amorphous and Crystalline Phases: The Case of TiO_2. *J. Phys. Chem. Lett.* **2018**, *9*, 2985–2990.

(54) Caravati, S.; Bernasconi, M.; Kühne, T. D.; Krack, M.; Parrinello, M. Coexistence of Tetrahedral- and Octahedral-like Sites in Amorphous Phase Change Materials. *Appl. Phys. Lett.* **2007**, *91*, 171906.

(55) Casiraghi, C.; Piazza, F.; Ferrari, A. C.; Grambole, D.; Robertson, J. Bonding in Hydrogenated Diamond-like Carbon by Raman Spectroscopy. *Diam. Relat. Mater.* **2005**, *14*, 1098–1102.

(56) Weiler, M.; Sattel, S.; Jung, K.; Ehrhardt, H.; Veerasamy, V. S.; Robertson, J. Highly Tetrahedral, Diamond-like Amorphous Hydrogenated Carbon Prepared from a Plasma Beam Source. *Appl. Phys. Lett.* **1994**, *64*, 2797–2799.

(57) Sque, S. J.; Jones, R.; Briddon, P. R. Structure, Electronics, and Interaction of Hydrogen and Oxygen on Diamond Surfaces. *Phys. Rev. B* **2006**, *73*, 085313.

(58) Moras, G.; Pastewka, L.; Gumbsch, P.; Moseler, M. Formation and Oxidation of Linear Carbon Chains and Their Role in the Wear of Carbon Materials. *Tribol. Lett.* **2011**, *44*, 355–365.

(59) Moras, G.; Pastewka, L.; Walter, M.; Schnagl, J.; Gumbsch, P.; Moseler, M. Progressive Shortening of sp-Hybridized Carbon Chains through Oxygen-Induced Cleavage. *J. Phys. Chem. C* **2011**, *115*, 24653–24661.

(60) Sainio, S.; Nordlund, D.; Caro, M. A.; Gandhiraman, R.; Koehne, J.; Wester, N.; Koskinen, J.; Meyyappan, M.; Laurila, T. Correlation between sp3-to-sp2 Ratio and Surface Oxygen Functionalities in Tetrahedral Amorphous Carbon (ta-C) Thin Film Electrodes and Implications of Their Electrochemical Properties. *J. Phys. Chem. C* **2016**, *120*, 8298–8304.

(61) Konicek, A. R.; Grierson, D. S.; Sumant, A. V.; Friedmann, T. A.; Sullivan, J. P.; Gilbert, P. U. P. A.; Sawyer, W. G.; Carpick, R. W. Influence of Surface Passivation on the Friction and Wear Behavior of Ultrananocrystalline Diamond and Tetrahedral Amorphous Carbon Thin Films. *Phys. Rev. B* **2012**, *85*, 155448.

(62) Eryilmaz, O. L.; Erdemir, A. TOF-SIMS and XPS Characterization of Diamond-like Carbon Films after Tests in Inert and Oxidizing Environments. *Wear* **2008**, *265*, 244–254.

(63) Bagri, A.; Mattevi, C.; Acik, M.; Chabal, Y. J.; Chhowalla, M.; Shenoy, V. B. Structural Evolution during the Reduction of Chemically Derived Graphene Oxide. *Nat. Chem.* **2010**, *2*, 581–587.

(64) Lin, L. C.; Grossman, J. C. Atomistic Understandings of Reduced Graphene Oxide as an Ultrathin-Film Nanoporous Membrane for Separations. *Nat. Commun.* **2015**, *6*, 8335.

(65) Mao, S.; Pu, H.; Chen, J. Graphene Oxide and Its Reduction: Modeling and Experimental Progress. *RSC Adv.* **2012**, *2*, 2643–2662.

(66) Caro, M. A.; Aarva, A.; Deringer, V. L.; Csányi, G.; Laurila, T. Reactivity of Amorphous Carbon Surfaces: Rationalizing the Role of Structural Motifs on Functionalization using Machine Learning. *Submitted*.

(67) Sosso, G. C.; Miceli, G.; Caravati, S.; Behler, J.; Bernasconi, M. Neural Network Interatomic Potential for the Phase Change Material GeTe. *Phys. Rev. B* **2012**, *85*, 174103.

(68) Bitzek, E.; Koskinen, P.; Gähler, F.; Moseler, M.; Gumbsch, P. Structural Relaxation Made Simple. *Phys. Rev. Lett.* **2006**, *97*, 170201.

(69) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092.

(70) Brenner, D. W.; Shenderova, O. A.; Harrison, J. A.; Stuart, S. J.; Ni, B.; Sinnott, S. B. A Second-Generation Reactive Empirical Bond Order (REBO) Potential Energy Expression for Hydrocarbons. *J. Phys. Condens. Matter* **2002**, *14*, 783–802.

(71) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and Accurate Density Functional Calculations Using a Mixed Gaussian and Plane Waves Approach. *Comput. Phys. Commun.* **2005**, *167*, 103–128.

(72) Kühne, T. D.; Krack, M.; Mohamed, F. R.; Parrinello, M. Efficient and Accurate Car-Parrinello-like Approach to Born-Oppenheimer Molecular Dynamics. *Phys. Rev. Lett.* **2007**, *98*, 066401.

(73) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; Van de Vondele, J. CP2K: Atomistic Simulations of Condensed Matter Systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 15–25.

(74) Goedecker, S.; Teter, M.; Hutter, J. Separable Dual-Space Gaussian Pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703–1710.

(75) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(76) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(77) Li, J. AtomEye: An Efficient Atomistic Configuration Viewer. *Model. Simul. Mater. Sci. Eng.* **2003**, *11*, 173–177.

(78) Stukowski, A. Visualization and Analysis of Atomistic Simulation Data with OVITO–the Open Visualization Tool. *Model. Simul. Mater. Sci. Eng.* **2010**, *18*, 015012.

(79) Momma, K.; Izumi, F. VESTA 3 for Three-Dimensional Visualization of Crystal, Volumetric and Morphology Data. *J. Appl. Crystallogr.* **2011**, *44*, 1272–1276.

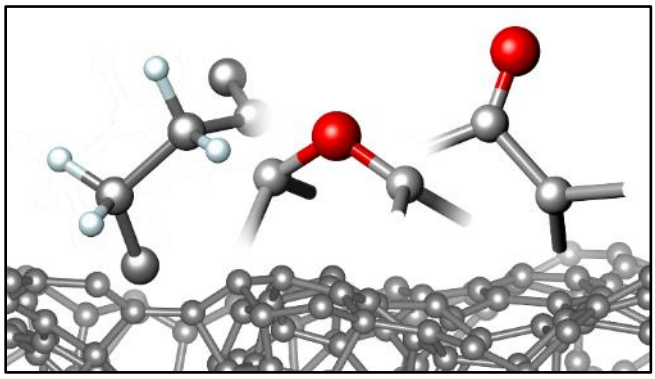

Table of Contents artwork

Supporting Information for the manuscript

Computational Surface Chemistry of Tetrahedral Amorphous Carbon by Combining Machine Learning and DFT

Volker L. Deringer, Miguel A. Caro, Richard Jana, Anja Aarva, Stephen R. Elliott, Tomi Laurila, Gábor Csányi, and Lars Pastewka*

*E-mail: vld24@cam.ac.uk

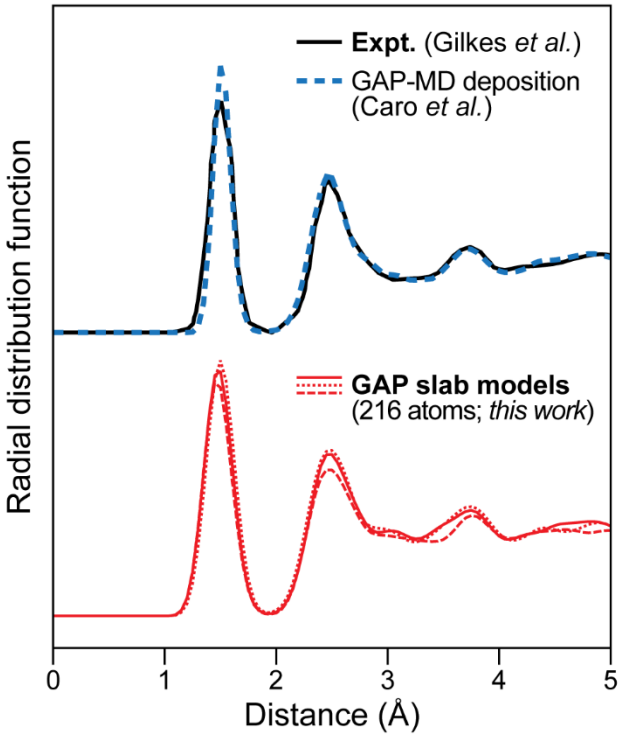

Figure S1. Radial distribution functions for *ta*-C systems (in arbitrary units and vertically offset for clarity), evidencing the ability of our slab models to describe the relevant features within the necessarily limited system size. The **upper part** is reproduced from recent work on large-scale, GAP-driven deposition simulations,[S1] where the RDF has been evaluated for the bulk-like center of the film "grown" in simulation; this serves as a benchmark for computational methods (dashed blue line).[S1] Also shown are neutron-scattering data taken from Gilkes *et al.*, providing an additional experimental benchmark (the authors give an atomic density of 1.50×10^{29} m^{-3}, corresponding to a mass density of 3.0 g cm^{-3}; solid line).[S2] The **lower part** of the panel gives results for the much smaller slab models created in this work. These data were obtained by excluding from the calculation the outermost 3 Å of the slab, which are deemed to be the "surface region" (see main text) and by applying Gaussian broadening of 0.1 Å to atomic positions. Three uncorrelated simulation cells have been arbitrarily chosen to exemplify the variation from system to system.

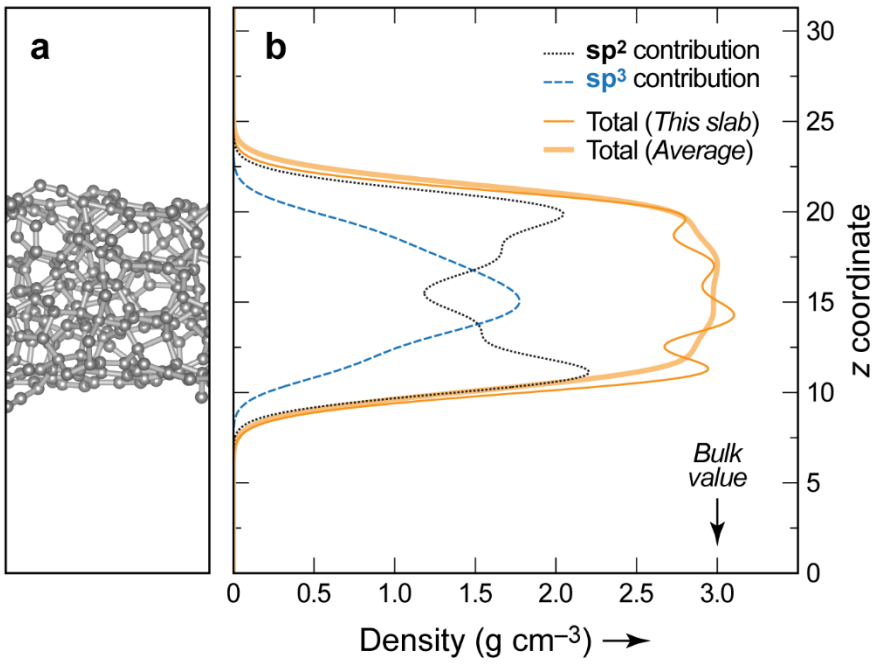

Figure S2. Density profiles, as in ref S1, calculated here for one of the 216-atom slab models (panel **a**). The atomic density is evaluated as a function of the coordinate normal to the surface (z), by applying Gaussian broadening to each atomic position. In panel **b**, a thin orange line shows the total density profile thus obtained for this particular structure; dashed black and blue lines indicate only sp^2 and only sp^3 contributions, respectively. An average of the total density profiles, collected over all 50 slabs, is also given (thick orange line). While the sp^2/sp^3 ratio changes notably near the surfaces, the slab center (at $z \approx 15$ Å) retains a relatively large sp^3 concentration (*cf.* Figure 2 in the main text), and a density close to the initially chosen bulk value of 3 g cm^{-3}. In line with the discussion in the main text, this provides further evidence that the simulation cells employed here can indeed provide a reasonable description of both bulk and surface regions.

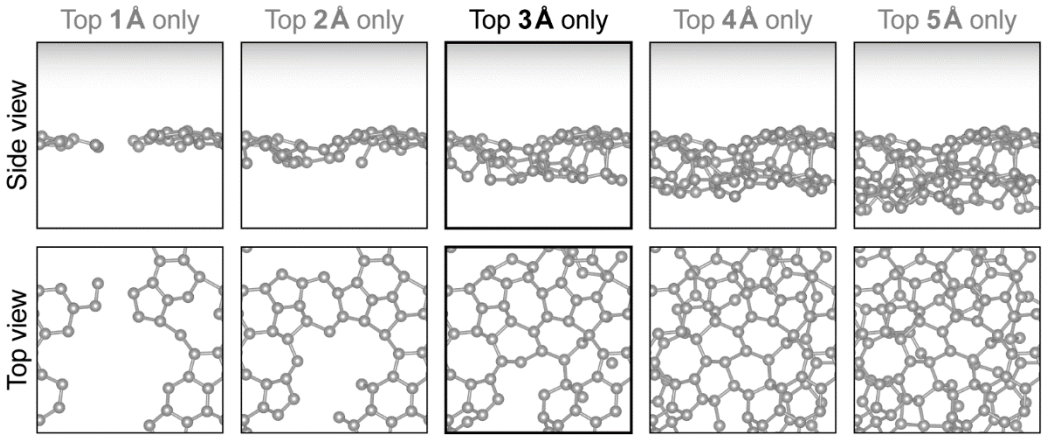

Figure S3. How large to choose the "surface region"? These structural drawings illustrate the choice of the topmost 3 Å as made throughout the main text. We show one exemplary 216-atom *ta*-C slab model in all panels, progressively including more atoms: from 1 Å to 5 Å below the topmost atom. Too narrow regions (*left-hand side*) fail to include substantial parts of the surface, whereas too large regions (*right-hand side*) include non-surface atoms. The setting of 3 Å (*middle panels*), while still necessarily being an ad-hoc choice, provides a good compromise.

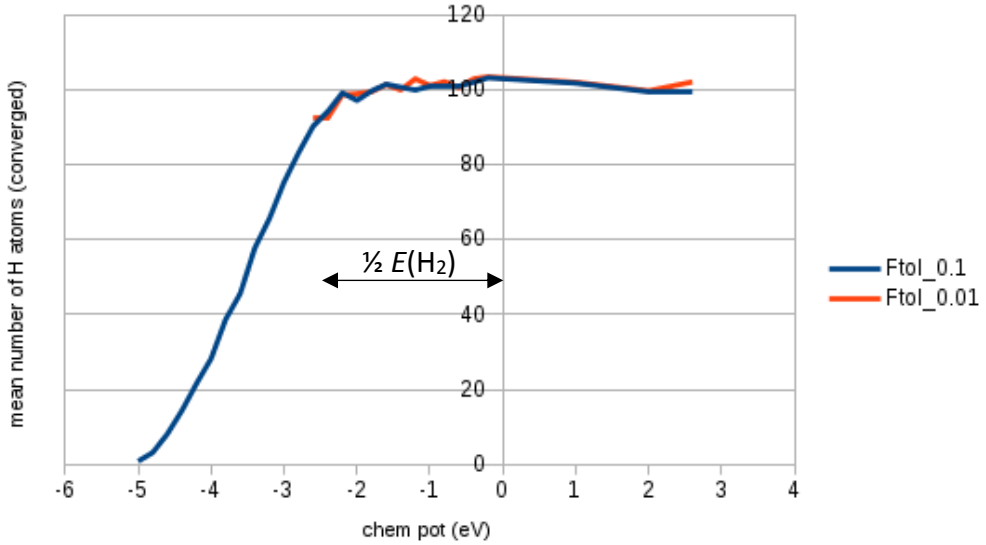

Figure S4 (to be re-plotted in final style). Results of several sets of independent Monte Carlo trajectories, starting from the same 50 slab models but with varying hydrogen chemical potential, μ. As a function of this potential, we evaluate the mean number of H atoms in the converged systems (corresponding to the *x*-axis in Figure 4 in the main text). This figure illustrates that the results are robust with respect to larger chemical potentials (≡ higher H abundance, indicating saturation). At lower chemical potentials, the H content begins to decrease because the acceptance term becomes less and less favorable. The chemical potentials can be linked to the partial pressure of gaseous species, and thus to so-called "*ab initio* thermodynamic" analyses as pioneered for crystalline surfaces in ref S3. This is an interesting direction of future work. For further validation, we also compare results with a strict force convergence criterion (0.01 eV $Å^{-1}$), as used in the main text (*red line*), to separate runs with a less strict criterion (0.1 eV $Å^{-1}$; *blue line*). This evidences that the results shown are well-converged with respect to this computational parameter, and it attests again to the robustness of the protocol.

Supplementary References

(S1) Caro, M. A.; Deringer, V. L.; Koskinen, J.; Laurila, T.; Csányi, G. Growth Mechanism and Origin of High sp^3 Content in Tetrahedral Amorphous Carbon. *Phys. Rev. Lett.* **2018**, *120*, 166101.

(S2) Gilkes, K. W. R.; Gaskell, P. H.; Robertson, J. Comparison of Neutron-Scattering Data for Tetrahedral Amorphous Carbon with Structural Models. *Phys. Rev. B* **1995**, *51*, 12303–12312.

(S3) Reuter, K.; Scheffler, M. Composition, structure, and stability of RuO_2(110) as a function of oxygen pressure. *Phys. Rev. B* **2001**, *65*, 035406.