Caro, Miguel A.; Aarva, Anja; Deringer, Volker L.; Csányi, Gábor; Laurila, Tomi

Reactivity of Amorphous Carbon Surfaces

Published in:
Chemistry of Materials

DOI:
10.1021/acs.chemmater.8b03353

Published: 01/01/2018

Document Version
Publisher's PDF, also known as Version of record

CHEMISTRY OF MATERIALS
Cite This: Chem. Mater. 2018, 30, 7446−7455
Article
pubs.acs.org/cm

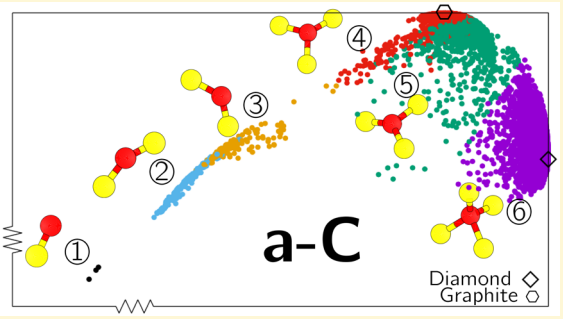
Reactivity of Amorphous Carbon Surfaces: Rationalizing the Role of Structural Motifs in Functionalization Using Machine Learning

Miguel A. Caro,*[,†,‡] Anja Aarva,[†] Volker L. Deringer,[§,∥] Gábor Csányi,[§] and Tomi Laurila[†]

[†]Department of Electrical Engineering and Automation, School of Electrical Engineering, Aalto University, Espoo 02150, Finland
[‡]QTF Centre of Excellence, Department of Applied Physics, Aalto University, Espoo 02150, Finland
[§]Engineering Laboratory, University of Cambridge, Cambridge CB2 1PZ, United Kingdom
[∥]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

Supporting Information

ABSTRACT: Systematic atomistic studies of surface reactivity for amorphous materials have not been possible in the past because of the complexity of these materials and the lack of the computer power necessary to draw representative statistics. With the emergence and popularization of machine learning (ML) approaches in materials science, systematic (and accurate) studies of the surface chemistry of disordered materials are now coming within reach. In this paper, we show how the reactivity of amorphous carbon (a-C) surfaces can be systematically quantified and understood by a combination of ML interatomic potentials, ML clustering techniques, and density functional theory calculations. This methodology allows us to process large amounts of atomic data to classify carbon atomic motifs on the basis of their geometry and quantify their reactivity toward hydrogen- and oxygen-containing functionalities. For instance, we identify subdivisions of sp and sp^2 motifs with markedly different reactivities. We therefore draw a comprehensive, both qualitative and quantitative, picture of the surface chemistry of a-C and its reactivity toward −H, −O, −OH, and −COOH. While this paper focuses on a-C surfaces, the presented methodology opens up a new systematic and general way to study the surface chemistry of amorphous and disordered materials.



I. INTRODUCTION

Understanding the surface chemistry of amorphous and disordered materials is a crucial step toward the rational design of cost-effective, tailor-made materials with targeted electrocatalytical properties. This has ramifications for the realms of biocompatible sensing applications,[1] nanoelectronics,[2,3] electrocatalysis,[4] and efficient energy generation, including renewable energy applications (photoelectrochemistry,[5] fuel cells,[6] CO_2 reduction,[7] etc.), just to name a few. The knowledge of specific interactions between the surface and analyte, including adsorption characteristics and atomic processes at the nanoscale, is often a missing piece in the wider puzzle of how material stoichiometry, growth process, surface morphology, and ultimate application performance are all connected to one another.

Amorphous carbon (a-C) is one such important disordered material. Specifically, dense sp^3-rich "tetrahedral" a-C (ta-C) and diamond-like carbon (DLC) have important scientific and industrial applications.[8] The mechanical properties of DLC, close to those of diamond, make it an ideal material to be used for coatings. The chemical properties of a-C, namely biocompatibility, chemical inertness, and resistance to corrosion and bacterial adhesion, have been at the root of recent interest in a-C as a substrate material for biological applications. In particular, biocompatible electrochemical sensors for *in vivo* analysis, where the electrode is coated with a-C, are of high topical and technological interest.[1] To predict how these electrodes interact with the analyte, a deep understanding of the surface chemistry of a-C is required.

In principle, a computational atomistic simulation would be an ideal approach to studying this material system. However, because of the disordered nature of a-C, systematic studies of adsorption characteristics and chemical reactivity must take into account the huge morphological and bonding variability exhibited by a-C. A successful attempt to tackle such a problem must necessarily rely on representative statistical sampling of the different atomic motifs encountered in realistic a-C surfaces. Because of the large number of structures to be considered, one needs to combine electronic structure methods, such as density functional theory (DFT), with automated tools to accelerate the calculations and to rationalize the results. The usefulness of such conceptual approaches extends way beyond the realm of amorphous carbon, being applicable to any disordered material.

Received: August 7, 2018
Revised: August 30, 2018
Published: September 10, 2018

DOI: 10.1021/acs.chemmater.8b03353
Chem. Mater. 2018, 30, 7446−7455

In this work, we seek a comprehensive understanding of the surface properties of a-C, combining DFT with machine learning (ML). We show how ML techniques can be used to rationalize the wealth of chemical and physical information that can be extracted from atomistic structure models and derive a new set of atomic and electronic descriptors that can efficiently predict adsorption energies (thereby quantifying chemical reactivity). For the first purpose, we use ML clustering techniques that allow us to classify atomic motifs and adsorption sites according to their geometrical features and correlate them with chemical reactivity toward different functional groups commonly found in a-C. Adsorption characteristics are then established by means of DFT calculations. We identify which a-C sites are most reactive toward chemisorption of hydrogen (−H), oxygen (−O), a hydroxyl group (−OH), and a carboxylic acid group (−COOH). These functional groups have been experimentally proven to be present on a-C surfaces[9] and play an important role in the surface chemistry of a-C and other disordered carbons, for instance, when these materials are employed as electrodes in electrochemical analysis.[9−11] Finally, we use these DFT values to train and optimize a ML model, based on the Gaussian approximation potential (GAP) framework, to predict adsorption energies from structural and electronic atomic descriptors. This demonstrates how a combined strategy of augmenting local structural features with local electronic descriptors can pave the way toward accurate adsorption models.

II. ATOMIC MOTIFS

II.A. Machine Learning-Based Structure Generation. In this work, we used a set of structural models that we generated in a preceding study.[12] Two-dimensional (2D) slabs were cleaved from extended structures by inserting an "artificial" vacuum normal to the surface, and then surface properties were studied by allowing for reconstruction, adding desired species, and so on. In ref 12, we used the a-C GAP that has been extensively validated with respect to structural and mechanical properties,[13] a correct description of the potential energy surface as probed by crystal structure searching,[14] surface energies,[13] and finally the description of the deposition process.[15] We systematically evaluated the system-size dependence of a-C slab modeling: one wants to use a model system that is as small as possible, in the interest of computational efficiency, which still needs to be large enough to provide a representative local structure. We found that 216 atoms per simulation cell are well suited for this task, corresponding to an in-plane length of ≈11 Å for the cell. The latter also defines the lateral spacing for adsorbate species. The surface slabs were generated by cleaving from bulk ta-C, heating to 1000 K over 10 ps in GAP molecular dynamics (MD), annealing at that temperature for 10 ps, and cooling back over an additional 20 ps. Details of these simulations and atomic coordinates of the pristine (chemically unmodified) simulation cells are provided in ref 12.

II.B. Clustering Algorithms. Having access to a large number of structures allows us to compute good statistics. In total, we have 10802 a-C atomic sites (including bulk diamond and graphite in the data set), all of which are strictly geometrically inequivalent. Making sense of and finding trends in such a large data set call for automated approaches and the use of artificial intelligence. There are two main tasks at hand here. One is to characterize each atomic site on the basis of its environment, preferentially in a chemically intuitive way. Another is to classify all of those sites so that similar sites are grouped together and trends in their properties, namely chemical reactivity, can be correlated with structure. For the first task, we use the smooth overlap of atomic positions (SOAP) approach,[16] which provides an intuitive measure of dissimilarity (or "distance", in the ML jargon) between atomic environments. SOAP is a new approach, increasingly used by the computational materials chemistry community, for "encoding" atomic environments into a numerical descriptor that can then be fed into ML algorithms.[17] The same method is used in GAP to compare atomic environments. In all cases, SOAP analysis is performed within a given "cutoff radius", which defines how far the SOAP algorithm "sees" the atomic environment; neighbors outside the cutoff will not affect the result. We find that a 2 Å cutoff radius, slightly larger than typical covalent bond lengths in the system, successfully captures both geometric variability and chemical trends. While larger cutoffs can be useful for ML models of, e.g., cohesive energies,[13] it then becomes difficult to visualize the motifs and make the connection with intuitive chemical concepts.

Once each atom has been assigned a SOAP vector with structural information, the distance/dissimilarity between environments is calculated as a dot product. In particular, we define the distance matrix element between environments i and j from the fourth power of the SOAP kernel:

$$D_{ij} = \sqrt{1 - [k(i, j)]^4} \tag{1}$$

where $k(i,j) = \mathbf{q}_i \cdot \mathbf{q}_j$, with $\mathbf{q}_i$ and $\mathbf{q}_j$ being the SOAP vectors that characterize the densities of sites i and j, respectively. D is square and symmetric. Our similarity matrix is simply given by

$$S_{ij} = [k(i, j)]^4 \tag{2}$$

All of these matrices have dimensions of $n \times n$, where n is the number of sites in the data set. In our case, n = 10802. Obviously, an understanding that can relate to chemical intuition must be built on reducing the dimensionality of this problem. The dimensionality of our problem is given by the total number of independent distances and/or similarities and equals n. That is, the coordinates of each point in the data set are characterized by its $n - 1$ distances to every other point in the data set (plus the self-distance, which is always zero). One could also choose to carry out the representation of the data set on the SOAP vector space, the dimensionality of which equals the number of components of the SOAP vectors; however, this does not resolve the issue because this representation would still be highly dimensional.

We propose to apply two different approaches, both reducing the dimensionality of the problem from n-dimensional to two-dimensional. One is to compute the similarity of each atomic site to diamond and graphite; this resonates with chemical intuition and establishes a strong link to the notions of sp^2- and sp^3-like chemical bonding. Another one is to use a ML technique called multidimensional scaling (MDS) that, in essence, projects the distances in the highly dimensional plane to a plane of reduced dimensionality (2D in this case) that optimally preserves the original distances. That is, the choice of a 2D plane is (iteratively) optimized such that the 2D distances in the new plane resemble the original $(n^2 - n)/2$ distances as accurately as possible. This approach allows us to simultaneously (i.e., on the same plot) visualize how different all the atomic sites are from one another. We use the MDS algorithm

DOI: 10.1021/acs.chemmater.8b03353
Chem. Mater. 2018, 30, 7446−7455

from the Python Scikit-learn library.[18] Each time the algorithm is run, it chooses a different random initialization. We run the algorithm 64 times and choose the solution that shows the lowest "stress", that is, the solution that provides the best 2D representation of our data set. SOAP descriptors have already been used in conjunction with visualization techniques to characterize differences between chemical environments.[19,20]

The classification of atomic sites is done using a ML "clustering" technique. Similar environments (atomic sites) belong to the same cluster. That is, intracluster distances D_{ij} and similarities S_{ij} must be small and large, respectively. To build the clusters, we use a variant[21] of *k*-medoids.[22] Our approach is flexible enough that it accepts a predefined target number of clusters (or atomic motifs) and does not introduce a bias due to some motifs being more frequent than others. Technical details about the cluster algorithm employed are given in the Supporting Information. All in all, we find that a 2 Å SOAP cutoff, together with a maximum of six clusters and the use of a "relative" intracluster coherence criterion, provides the best recipe in terms of classifying atomic motifs in a-C in accordance with chemical intuition (as will be shown next). The remainder of this work will adopt this as convention.

II.C. Motif Identification and Cataloging, Bulk and Surface. The results of the clustering analysis, for the 50 different a-C slabs, graphite, and diamond, that is, a total of 10802 sites, are shown in Figure 1. In Figure 1, we plot the

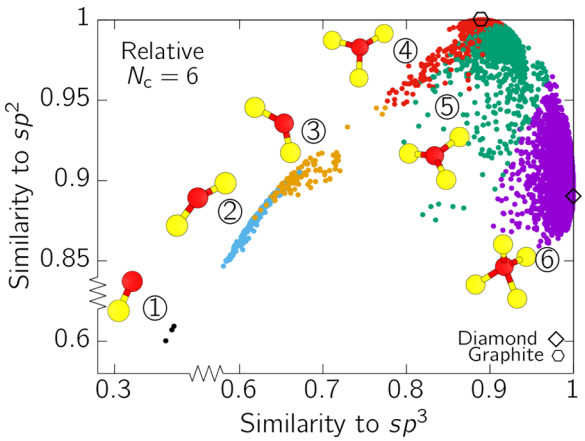


Figure 1. Results of the clustering analysis with six target clusters and the relative coherence criterion. Atomic sites that belong to the same cluster are represented with dots of the same color. Results for different criteria are shown in the Supporting Information. Overlaid on the graph is a ball and stick representation of the medoid of each cluster. Red atoms represent the atomic sites in question, and yellow atoms represent its nearest neighbors.

position of each atomic environment relative to its similarity to diamond (sp^3) and graphite (sp^2), based on a 2 Å SOAP cutoff. While the standard way of assigning sp^2 and sp^3 character, commonplace in the literature, relies on simply counting neighbors, our approach takes the detailed atomic structure into account via the SOAP descriptors. In Figure 1, the clusters are numbered systematically by increasing coordination. Cluster 1 corresponds to C sites with only one neighbor that are therefore coordination defects (only three samples of 10802 in the data set). Cluster 6 comprises sp^3-like sites, which are similar to diamond, with four atomic neighbors. Ball-and-stick representations of the medoids for all these clusters are shown in Figure 1. Coordinates for these medoids are provided in the Supporting Information. We can observe that the medoids corresponding to clusters 2 and 3, on one hand, and 4 and 5, on the other, are very similar to each other. The differences between them are primarily due to bond angle bending, for sp-like sites, and bond distances, for sp^2-like sites, as evidenced by the histogram in Figure 2, where we

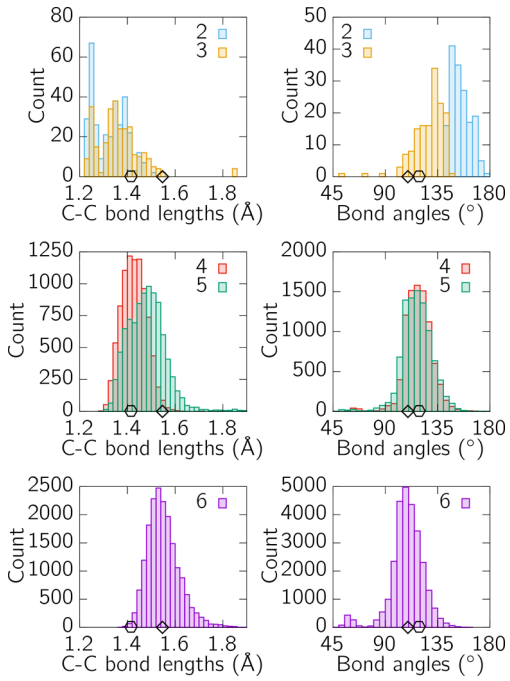


Figure 2. Distribution of bond lengths and bond angles for the different variants of the identified a-C atomic motifs. Rhombi (◇) and hexagons (⬡) denote the diamond and graphite values, respectively.

show the distributions of bond distances and angles. In the Supporting Information, we further show motif nonlinearity and nonplanarity *h* for sp and sp^2 sites, respectively. Figure 2 reveals that the main difference between sites belonging to clusters 2 and 3 (sp) is the bond angle, distributed around 155° and 130°, respectively. For the two kinds of sp^2 motifs recognized by the algorithm, we observe a homogeneous distribution of angles around 120°, which is the ideal graphite value. The main difference is the shorter average bond length for cluster 4, around 1.42 Å, compared to ∼1.5 Å for cluster 5. The values for cluster 4 are also significantly more narrowly distributed. For these sp^2 motifs, we observe that the SOAP analysis tends to emphasize more radial density differences than angular density differences. The importance of bond directionality is highly system-dependent. Typically, ionic bond character and covalent bond character emphasize bond distances and bond angles, respectively, as highlighted by a detailed study of internal strain (Kleinmann parameter) in tetrahedrally bonded III−V semiconductors.[23] Extending the SOAP formalism to separately weight the importance of bond angles and bond distances will provide improved flexibility and accuracy of future GAP models.

To gain insight into surface reactivity, we look in more detail at surface sites. Because a "surface region" is usually defined in a somewhat arbitrary manner, here we use a probe-sphere algorithm as implemented in CCP4's AREAIMOL tool[24,25] to identify surface sites. The used van der Waals and probe-

DOI: 10.1021/acs.chemmater.8b03353
Chem. Mater. 2018, 30, 7446−7455

sphere radii are 1.8 and 2 Å, respectively. Surface and interior ("bulk") atoms in the a-C slabs are identified in this fashion. In Figure 3, we therefore extend the analysis of Figure 1 by

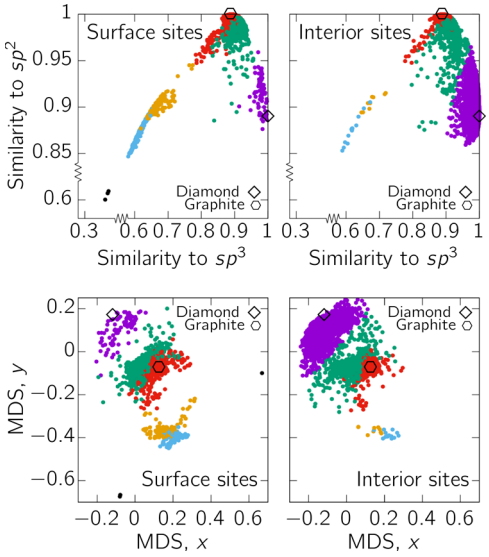


Figure 3. Maps of atomic sites separated into bulk (interior of the slab) and surface sites. The top panels show a representation based on similarity to sp^2 and sp^3, and the bottom panels show a 2D representation, (x, y), based on MDS dimensionality reduction.

separating between bulk and surface sites and using both the sp^2/sp^3 plotting method and the dimensionality reduction scheme [multidimensional scaling (MDS)] outlined in section II.B. Diamond and graphite are highlighted on the plots, as a guide. As expected,[15] we observe that surface and bulk sites are distributed differently. While sp^2-like motifs can be found in both the surface and interior of the slabs, sp and sp^3 sites are found predominantly only in the surface and interior, respectively. The MDS approach places graphite right in the middle of cluster 4 that, as seen in Figure 2, shows bond lengths closer to those of graphite than those of the other sp^2-like cluster (cluster 5). On the other hand, diamond is placed by this scheme in the periphery of the cluster of sp^3-like motifs. Because we have not introduced any intuitive bias into the scheme, MDS is a useful guide for motif classification. For instance, it confirms graphite as a good exemplary sp^2 motif but tells us that diamond is not a good example of a 4-fold coordinated site, because it lies far from the middle of cluster 6. On the basis of these observations, we speculate that MDS representation could help in the classification and identification of motifs in other amorphous materials, such as a-Si, phosphorus, etc.

III. SURFACE REACTIVITY

To explore and understand the chemical reactivity of the a-C surfaces, we calculate adsorption energies for a set of functional groups on different adsorption sites, as identified in the previous section.

III.A. Adsorption Energy Calculations. Adsorption energies are obtained as the difference between the total energy of the whole system (slab plus adsorbed group) E_{tot} and the sum of the total energies of the slab (E_{slab}) and the isolated group in vacuum (for H, E_H). Therefore, more negative energies correspond to more favorable adsorption. Total energies are calculated within the framework of DFT with projector augmented-wave (PAW) potentials,[26,27] as implemented in GPAW.[28,29] We use the Perdew–Burke–Ernzerhof (PBE) exchange-correlation density functional.[30] van der Waals (vdW) corrections are applied via the method introduced by Tkatchenko and Scheffler.[31] Reciprocal space is sampled using a Monkhorst–Pack (MP) grid[32] with $2 \times 2 \times 1$ k-point sampling. Because amorphous or defected carbonaceous materials are known to possess local (atomic) magnetic moments,[1,33] all calculations are performed with spin polarization. The GAP-generated slabs had been previously relaxed with a different DFT code (without vdW corrections, but using the same functional, viz. PBE) by Deringer et al.[12] using spin-paired calculations. To ensure the optimal accuracy of the adsorption energy calculations, we further relax the geometry of the slabs with GPAW using spin polarization and including vdW corrections.

III.B. Probing Site Reactivity with Hydrogen. To obtain a measure of surface reactivity that can be assigned to the different motifs in a statistically significant manner, we conduct restricted-geometry adsorption calculations for H. Essentially, a H atom is placed 1.1 Å from the surface atomic site of interest, in a position that maximizes its distance to that site's nearest neighbors. The energy difference between structures before (slab and H separated) and after placing the H is plotted. The atoms are not allowed to relax during this test (the effect of full adsorption, including geometry optimization, will be studied in the next section). The distance maximization with respect to the site's nearest neighbors is based on a penalty function:

$$\sum_{i \in \mathrm{NN}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_H(\theta, \phi)|}, \quad \text{s.t.} |\mathbf{r}_H| = 1.1\ \text{Å} \tag{3}$$

That is, the H atom is placed at the position, away from the central atom, that minimizes the penalty function above, subject to the condition that the distance between the H atom and the central motif is constant and equal to 1.1 Å. The summation is performed over the nearest neighbors of the central motif. This problem is easiest to solve in spherical coordinates, where the optimization is performed directly with respect to angles θ and ϕ without needing to explicitly enforce the constraint. We refer to this approach as "hydrogen probing" for site reactivity. The results of our analysis are shown in Figure 4. Unsurprisingly, sp^3 sites are the most stable (some showing positive adsorption energies). Both sp^2 motifs (i.e., clusters 4 and 5) are similarly reactive, with cluster 5 being on average slightly more reactive (more negative adsorption energies). This resonates with our intuitive expectation, drawn from Figure 2, that, for a given coordination, the motifs with longer average bonds (cluster 5) will be less stable than motifs with shorter bonds (cluster 4). sp motifs show large variability in adsorption characteristics. The more bent motifs in cluster 3 are significantly more reactive than the flatter motifs in cluster 2 (this can also be observed in the local density of states in the next section). This is a success of our new classification scheme: solely on the basis of geometrical information, motifs that are seemingly very similar (according to coordination, sp sites in this case) have been classified separately into two classes with markedly different reactivity. The results of this section are summarized in Table 1.

III.C. Functionalization. To probe reactivity of a-C surfaces under more realistic conditions, we perform a series

DOI: 10.1021/acs.chemmater.8b03353
Chem. Mater. 2018, 30, 7446–7455

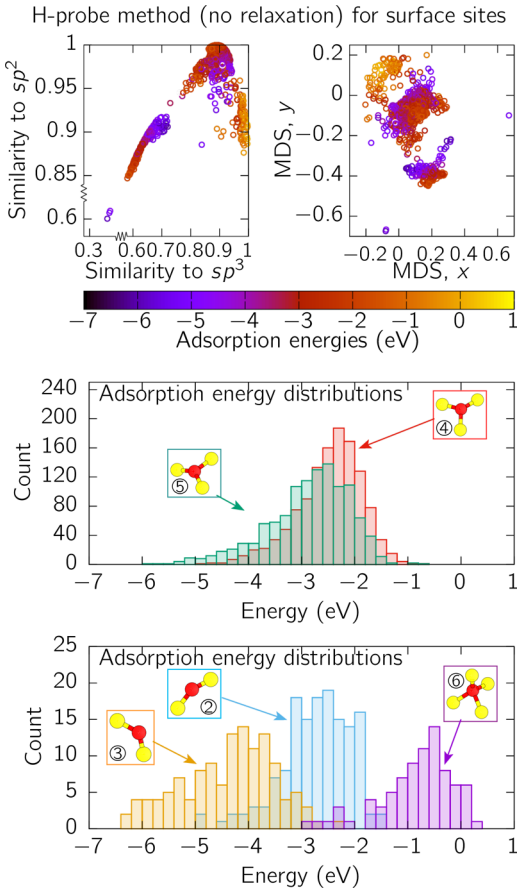


Figure 4. Results of H-probe analysis. (Top) Scatter plots of adsorption energies as a function of geometrical features and (bottom) distribution of adsorption energies for the different identified clusters.

Table 1. Summary of the Average Values from Figure 2 (geometrical) and Figure 4 (reactivity), with Standard Deviations, from Most Reactive to Least Reactive (according to the H-probe method)

cluster	description	$\overline{d}_{CC}$ (Å)	$\overline{\theta}_{CC}$ (deg)	$\overline{E}_{ad}^{H}$ (eV)
3	bent sp	1.365 ± 0.096	128 ± 13	−4.15 ± 1.97
5	long sp^2	1.481 ± 0.078	117 ± 13	−2.80 ± 1.21
2	straight sp	1.325 ± 0.069	155 ± 8	−2.73 ± 0.65
4	short sp^2	1.429 ± 0.053	118 ± 12	−2.42 ± 1.00
6	sp^3	1.551 ± 0.066	109 ± 14	−0.83 ± 0.69

of adsorption energy calculations for the functional groups expected to be most abundant at a-C surfaces.[9] Selected sites on these surfaces are functionalized with either hydrogen (−H), oxygen (−O), hydroxyl (−OH), or carboxylic acid groups (−COOH), and the geometry of the system is allowed to relax (see Figure 5). The adsorption sites for the groups are chosen according to the SOAP-based clustering scheme presented above. This way, we can try to establish a connection between the structural trends of the different adsorption sites and their reactivity. To compare the chemistry and binding properties of these sites, adsorption energies of all functional groups considered are computed for each cluster. The simulations are performed following the methodology outlined in section III.A.

Examples of the geometries of the sites in each cluster discussed in this work are depicted in Figure 1. Some motifs appear more frequently than others, and this is reflected in the number of elements in each cluster. For instance, among all the slabs considered in this study (>10000 total sites), only three sites belong to cluster 1, which consists of a C motif with only one neighbor. All other motifs appear more frequently, and thus, we can draw better adsorption statistics for those. Clusters 2 and 3 contain sp motifs, typically contained along a carbon chain that forms a ring on the surface. Clusters 4 and 5 contain sp^2 motifs. Cluster 6 corresponds to sp^3 sites. Given the computational cost of these simulations, only a limited number of adsorption sites (∼20 per cluster) are selected. The exceptions are cluster 1, for which we have only three sites, as discussed, and cluster 6, which shows extremely poor adsorption, due to its sp^3 nature, and is excluded from the study. The adsorption sites are chosen to be closest to the medoid of the cluster to which they belong (where "close" carries the meaning of distance discussed in section II.B).

The distributions of calculated adsorption energies are presented in Figure 6. Cluster 1 (C motif with only one neighbor) contains only three sites, and thus, the sampling is too poor to draw statistics. Sites in cluster 6 (sp^3 sites) do not favor adsorption, and bond breaking in the carbon matrix surrounding the adsorption site occurs every time a functional group is placed nearby. Occasionally, bond breaking occurs also for other clusters. For O adsorption, bond breaking in the a-C slab happens ∼15 and ∼20% of time for sp^2 and sp adsorption sites, respectively. For sp^2 sites, this O adsorption-induced bond breaking in the carbon matrix is also accompanied by ether formation (the oxygen atom is shared by two carbons that are not bonded to each other). These numbers are consistent with our previous observations.[12] Whenever bond breaking takes place, the adsorption site no longer represents the original motif. Because we are interested here in the reactivity of the original motif, adsorption energies on sites for which bond breaking occurs are not presented. Because of these considerations, results for clusters 1 and 6 are not included in Figure 6. We observe, for clusters 2 and 3 (sp motifs) and clusters 4 and 5 (sp^2 motifs), that sites that belong to different clusters display markedly different adsorption energies. We find sp motifs to be more reactive than sp^2 motifs. The largest differences in adsorption energies range between ∼2 eV more negative (H adsorption) and ∼3.5 eV more negative (O adsorption), for cluster 3 (sp) compared to cluster 4 (sp^2). While the reactivity of the different adsorption sites

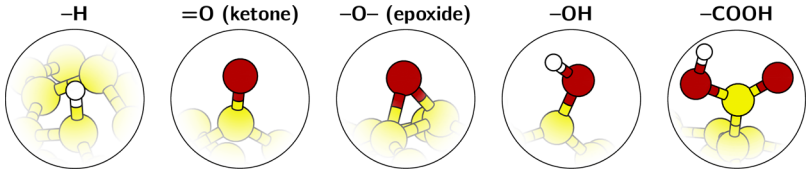


Figure 5. Functional groups explored in this study. Carbon, oxygen, and hydrogen atoms are colored yellow, dark red, and white, respectively.

DOI: 10.1021/acs.chemmater.8b03353
Chem. Mater. 2018, 30, 7446−7455

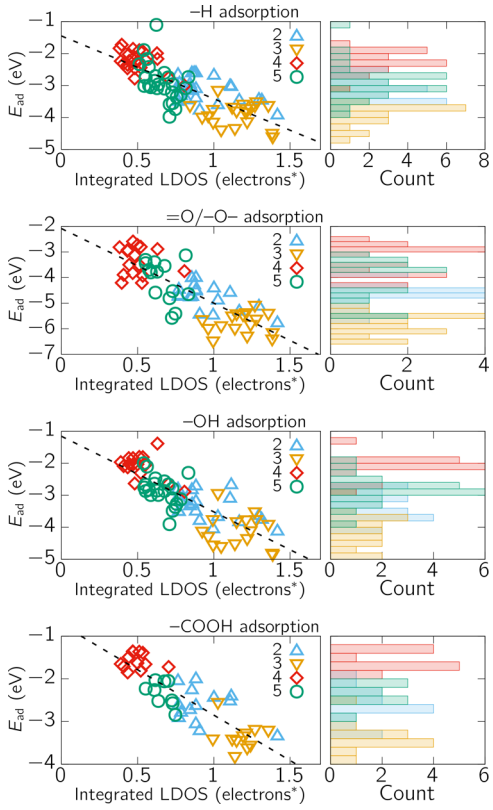


Figure 6. Adsorption energies (E_{ad}) of the functional groups vs the integrated local density of states (LDOS) for each site in each cluster, for clusters 2 and 3 (sp) and clusters 4 and 5 (sp^2). Dashed lines are linear fits to the data. Note that the integral of the LDOS equals the corresponding number of electrons only if the local basis used for the DOS projection is complete. We use atomic orbitals, which do not form a complete basis and lack full representation especially of the conduction band states. However, these integrated LDOS values should be a good guide for the actual (complete basis limit) relative ordering.

toward −H and −OH groups is similar, −O adsorption shows a stark increase in adsorption energies, with some sites showing adsorption energies as large as −6.5 eV. The interaction of the different motifs with the −COOH group is the weakest among the tested functionalizations. In all cases, the ordering of adsorption energies is the same and is consistent with the H-probe results.

To gain further insight into the connection among geometrical features, electronic structure, and reactivity, in Figure 6 adsorption energies are plotted versus the local density of states (LDOS) integrated around the Fermi level. Occupied electronic states below the Fermi level are weakly bound, and empty states above the Fermi level can easily accept electrons. Therefore, these states will be involved in chemisorption of functional groups, and the number of states (as given by the integrated LDOS) can act as a potentially good descriptor for site reactivity. The interval that is used for integration is from −3 to 3 eV. The LDOS integrated within this interval shows the best correlation with adsorption energies. The average LDOS for each cluster is depicted in the Supporting Information. The higher the density of states around the Fermi level, the more reactive the site in question is expected to be. In a similar way, transition metal d-band occupation has previously been shown to determine the characteristics of hydrogen chemisorption and used to rationalize trends in electrocatalysis.[34,35]

From Figure 6, we see that the integrated LDOS values correlate strongly with the adsorption energies. Figure 6 clearly shows that, when the LDOS around the Fermi level is high, adsorption energies are more negative, and vice versa. Furthermore, sites in a certain cluster are gathered around similar adsorption energy values. Indeed, while the general relation between LDOS and adsorption energy is clear, the specific correlation between them is heavily cluster-dependent. This is strong evidence that the clustering technique used here allows one to link motif geometry and adsorption energetics of a-C surfaces in a robust manner. Therefore, while geometrical features (clustering) as descriptor offers better performance than LDOS, combining the two, one could fit a ML model that could accurately predict the adsorption energies of a-C surfaces without the need to explicitly run the DFT calculation. We will deal with precisely this issue in the next section. Cluster 4 (sp^2 motif) seems to display the weakest interaction with the functional groups studied, with the exemption of cluster 6 (sp^3 motif), which is not shown in the figure. Cluster 1 is also missing from this analysis, because the sampling size is very small, comprising only three sites. We verify (not shown) that cluster 1 sites present the most negative adsorption energies of all the motifs studied. This is unsurprising because the sites in cluster 1 are coordination defects: they are so reactive that, under experimental conditions, they would be instantly terminated with any reactive species within interaction distance from the site or even already during deposition.

Figure 6 shows that the interaction between −O and the a-C surface is more complicated than the interaction between a-C and −H, −OH, and −COOH. In the case of oxygen, the adsorption energies are more scattered, both overall and within each cluster. The behavior of oxygen is different from those of the other groups because oxygen can become bonded to the C site in various ways. From our fully relaxed adsorption calculations, we observe that oxygen tends to form mostly either ketone or epoxide types of bonds. That is, the oxygen atom binds to one carbon with a double bond or becomes shared between two carbon atoms, respectively. Oxygen can also relax as an ether or a structural intermediate between an ether and an epoxide, although we observe only a few of these groups. This indicates that classical specification of the bond types (used widely in organic chemistry, for instance) does not fully apply in the case of a-C and oxygen, as evidenced by our DFT results. Indeed, in this context, the nature of bonding between a-C and −O seems to be difficult to describe in classical terms.

We summarize all the results of our study of functionalization (geometrical features and adsorption energies) in Table 2. Average values are shown, together with standard deviations, for each combination of a motif (cluster) and a functional group that we have explored. It is manifest, in all cases, that when adsorption energies become more negative bond lengths become shorter, as expected. Another expected trend is that when the hybridization of the site changes via introduction of the adsorbant from sp^2 to sp^3 and from sp to sp^2, the bond angles approach 109° and 120°, respectively. In the case of epoxide groups, oxygen is bonded to two carbons that are in turn bonded to each other (cf. Figure 5). The fact that epoxides appear less often than ketones can be explained by ring strain arising from the carbons being forced into an approximately ∼60° bond angle. This makes the structure

DOI: 10.1021/acs.chemmater.8b03353
Chem. Mater. 2018, 30, 7446−7455

Table 2. Geometries and Energetics of the Different Functionalizations of a-C Surfaces Explored in This Work[a]

−H				
cluster	N	d_{HC} (Å)	θ_{HC} (deg)	E_{ad} (eV)
1	3	1.074 ± 0.005	168 ± 17	−4.48 ± 0.64
2	24	1.097 ± 0.002	118 ± 2	−3.15 ± 0.38
3	20	1.094 ± 0.002	119 ± 2	−3.90 ± 0.37
4	21	1.110 ± 0.004	107 ± 1	−2.24 ± 0.33
5	27	1.103 ± 0.006	108 ± 2	−2.89 ± 0.59
=O (ketone)				
cluster	N	d_{OC} (Å)	θ_{OC} (deg)	E_{ad} (eV)
1	3	1.184 ± 0.012	176 ± 3	−7.62 ± 0.38
2	18	1.250 ± 0.032	121 ± 5	−4.74 ± 0.50
3	16	1.232 ± 0.008	122 ± 2	−5.80 ± 0.41
4	9	1.346 ± 0.009	110 ± 1	−2.95 ± 0.33
5	10	1.307 ± 0.016	112 ± 2	−3.74 ± 0.53
−O− (epoxide)				
cluster	N	d_{OC} (Å)	θ_{COC} (deg)	E_{ad} (eV)
4	7	1.447 ± 0.009	62 ± 2	−3.87 ± 0.29
−OH				
cluster	N	d_{OC} (Å)	θ_{OC} (deg)	E_{ad} (eV)
1	2	1.302 ± 0.007	173 ± 3	−4.69 ± 0.45
2	20	1.375 ± 0.010	118 ± 2	−3.36 ± 0.43
3	16	1.363 ± 0.010	119 ± 2	−4.12 ± 0.51
4	16	1.449 ± 0.008	109 ± 1	−2.09 ± 0.37
5	22	1.424 ± 0.011	110 ± 1	−2.85 ± 0.44
−COOH				
cluster	N	d_{CC} (Å)	θ_{CC} (deg)	E_{ad} (eV)
2	14	1.505 ± 0.010	118 ± 1	−2.62 ± 0.40
3	11	1.494 ± 0.012	120 ± 1	−3.36 ± 0.32
4	12	1.568 ± 0.007	109 ± 1	−1.58 ± 0.18
5	10	1.545 ± 0.028	109 ± 2	−2.21 ± 0.49

[a]We show average values and their standard deviations. N is the number of sites sampled per each combination of a cluster and a functional group. For the epoxide groups, further geometrical values are as follows: d_{CC} = 1.500 ± 0.036 Å, and θ_{OCC} = 59 ± 1°.

unstable. In the table, we focus on the bond lengths and angles between the functional groups and the carbon matrix. The internal geometrical parameters of the −OH and −COOH groups show a very weak dependence on the adsorption site in question.

These data provide a quantitative complement to the trends that can be visualized throughout the figures in this section. We note that these numbers, although obtained for a-C surfaces, should be representative of typical values in carbon nanostructures. Our results should be particularly transferable to other disordered forms of carbon where passivation with oxygen- and hydrogen-containing functional groups is prevalent, such as graphene oxide,[36] reduced graphene oxide, and diamond.[37]

IV. PREDICTIVE POWER OF ML-BASED ADSORPTION MODELS

In the preceding sections, we have explored in detail the observed statistical properties of a-C atomic motifs, in terms of geometrical features, LDOS, and adsorption energies. We have also established the correlation between adsorption energies for different functional groups and, separately, a site's geometry and integrated LDOS. In this section, we go one step further and explore the ability of a ML model to predict the adsorption energies on an atomic site from a combination of atomic descriptors. In particular, we look at using geometry only via SOAP descriptors and enhancing SOAP with LDOS information. A model with good predictive ability will be a useful tool for estimating the degree of functionalization induced once a pristine a-C surface is placed in contact with some reactive environment, e.g., a regular atmosphere or an electrolyte. Understanding the connection between surface chemistry and catalytical/electrocatalytical performance will enable the development of tailored functional materials for specific purposes in energy applications, biosensing, the chemical industry, etc.

IV.A. ML Model and Kernel Optimization. Our ML model for adsorption energy prediction is a GAP model, described in detail in refs 38 and 39. Very briefly, an adsorption energy on site i is interpolated as follows:

$$E_{ad}^{i} = \sum_{t=1}^{N_t} \alpha_t k(i, t) \tag{4}$$

where t runs through all N_t configurations in the training set, α_t values are the fitting coefficients, and $k(i,t)$ is the similarity measure, or kernel, between site i and site t in the training set. The ability of this model to yield satisfactory predictions lies, to a great degree, in the choice of a suitable kernel. This kind of interpolation is much more sensitive to the choice of kernel than, for instance, the classification made in section II.A, where we focus on local chemical structure only.

Here, we introduce a new kernel that takes both atomic and electronic structure into account. We show that this kernel outperforms a purely structural approach in the fitting and prediction of adsorption energies. The first component of our kernel is based on SOAP descriptors $\mathbf{q}$ with varying cutoff r_c, as already described in section II.A:

$$k_1(i, j) = [\mathbf{q}_i(r_c)\cdot\mathbf{q}_j(r_c)]^{\zeta} \tag{5}$$

where ζ is some exponent, e.g., ζ = 4 in eq 1. $k_1(i,j)$ accounts for geometrical similarities only. The other kernel component is based on augmenting $k_1(i,j)$ by adding LDOS information. Because the LDOS is a continuous variable, we seek a compact (discrete) representation by computing its moments. The nth moment of the LDOS, computed in the vicinity of the Fermi level, is given by

$$\mu_n(i) = \int_{E_F-\Delta}^{E_F+\Delta} \mathrm{d}E(E - E_F)^n \mathrm{LDOS}^{(i)}(E) \tag{6}$$

where we choose Δ = 3 eV. These moments allow us to represent the LDOS in a manner similar to how a multipole expansion is used to represent a charge distribution. Using the LDOS moments allows us to construct the following kernel based on Gaussian distributions:

$$k_2(i, j) = k_1(i, j) \prod_{n=0}^{n_{max}} \exp\left\{-\frac{1}{2}\frac{[\mu_n(i) - \mu_n(j)]^2}{\sigma_n^2}\right\} \tag{7}$$

where σ_n controls how distant the nth LDOS moments of sites i and j can be to be considered "similar". For the models presented here, we compute up to the fifth moment (n_{max} = 5). The idea of constructing a SOAP+LDOS kernel is schematically depicted in Figure 7a.

DOI: 10.1021/acs.chemmater.8b03353
Chem. Mater. 2018, 30, 7446−7455

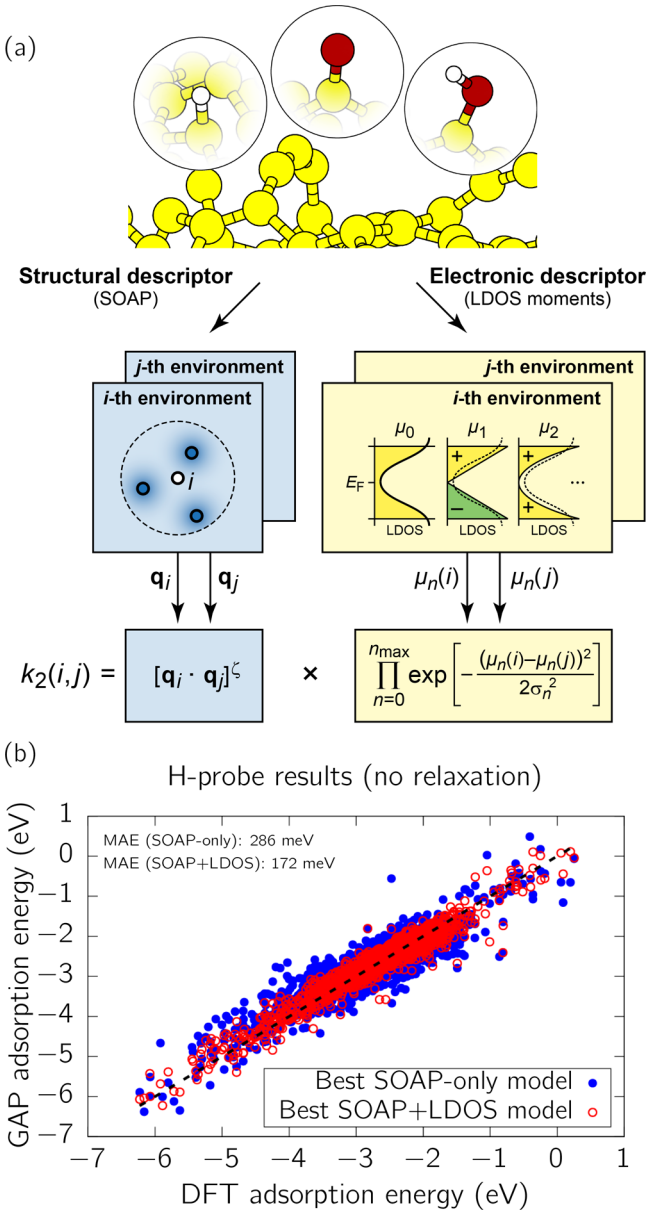


Figure 7. (a) Schematic view of the idea of constructing a SOAP +LDOS kernel. (b) Comparison of best SOAP-only and SOAP +LDOS GAP models.

The SOAP-only kernel has four parameters to be optimized, including the mentioned r_c and ζ. The SOAP+LDOS kernel has six additional parameters, the σ_n, for a total of 10. The number of training configurations in the set can also be added as a parameter of the overall ML model. We have optimized these parameters, using Monte Carlo sampling, by training and testing a total of ∼300k GAP ML models on the H-probe data (half used for training and half for testing). More details about this procedure are given in the Supporting Information. The "best" models are obtained by minimizing the root-mean-square error (RMSE) of the test set, which is an effective way of reducing the error due to outliers. Refinement of the model using conjugate gradient minimization from the best Monte Carlo result yields very marginal improvement (∼1 meV), which is a sign that the Monte Carlo procedure works almost optimally for this problem. Interestingly, while the optimal cutoff radius for the SOAP-only kernel (r_c) is 2.9 Å, this value is reduced for the SOAP+LDOS kernel to 2.3 Å.

The performance of the best (of 40k) SOAP-only model and the best (of 200k) SOAP+LDOS model is shown in Figure 7b. The RMSE's for predicted (GAP) versus measured (DFT) adsorption energies are 373 and 228 meV for SOAP-only and SOAP+LDOS models, respectively. The mean absolute errors (MAE's) are 286 and 172 meV, respectively. Therefore, inclusion of LDOS information allows us to significantly improve the prediction power of this model, reducing the error by ∼40%.

We note that computing LDOS still requires a DFT calculation. However, at least two reasons make a SOAP +LDOS model extremely useful. One is that for a supercell with N adsorption sites, probing all the adsorption energies directly would involve $O(N)$ full geometry optimizations or path calculations with DFT, thus potentially hundreds or thousands of additional DFT calculations. In contrast, LDOS for all N sites prior to adsorption can be computed with one single DFT calculation. The second reason is that an extremely precise representation of the LDOS may not be required, because in our model only the LDOS moments are taken into account (thus neglecting the fine detail of the LDOS). This means that a cheap DFT calculation with relaxed convergence parameters may be enough. We speculate that perhaps even a tight-binding LDOS calculation could be used to evaluate this new kernel.

IV.B. Prediction of Adsorption Energies for Different Functionalizations. Having optimized our kernel with the wealth of data available from the H-probe simulations, we now use the optimized parameters to train GAP models for interpolation of the adsorption energies of the different a-C functionalizations explored in section III.C. Because those data sets are much smaller than the H-probe one, the kernel parameters cannot be directly optimized with them. Again, because these data sets are so small ($50 \lesssim N_t \lesssim 100$) the training and testing is done in a different way, using N-fold cross validation in this case. The performance of our models, including MAE and RMSE for each model, is summarized in Figure 8 and Table 3. The results show a remarkable transferability for the kernel between the data set from which it was optimized (H-probe results) and these full adsorption estimates, considering the limited amount of data available to fit the model. In all cases, the global errors, listed in Table 3, are dominated by a few outliers. As in the previous section, we have omitted the undercoordinated ("one-fold") sites in cluster 1, which are discussed in the Supporting Information.

In all cases, the scatter of data is greatly reduced compared to that of the linear regression curve showed in Figure 6, which is essentially equivalent to a GAP model using the zeroth moment of the LDOS μ_0 as the sole descriptor (i.e., also excluding the geometrical information encoded in the SOAP). Unsurprisingly, O adsorption shows the worst results, where the error is dominated by a few outliers. Because of the more complex adsorption chemistry of O on C, building a ML model that can simultaneously predict adsorption energies for O atoms bonded to two carbon neighbors and one carbon neighbor requires further work. Such a model must be built on significantly more O adsorption data and may require further kernel optimization. Future work will deal with refinement and extension of these ML models to more general situations. The presented results open the door for accurate ML-based adsorption models that will become useful for predicting the

DOI: 10.1021/acs.chemmater.8b03353
Chem. Mater. 2018, 30, 7446−7455

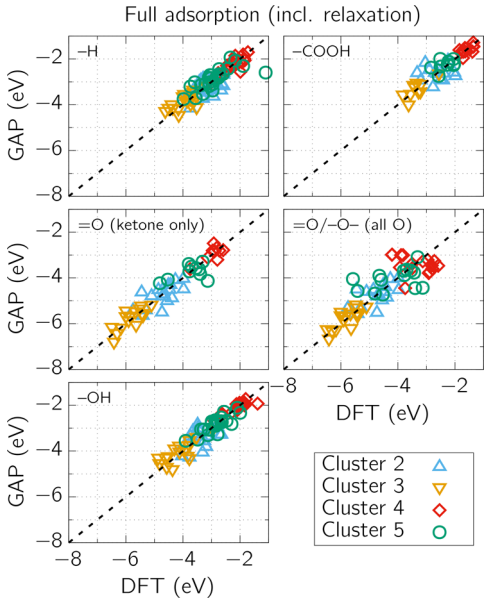


Figure 8. SOAP+LDOS GAP models for adsorption energy prediction on a-C surface sites.

Table 3. Performance (error estimates) of the GAP ML Models for Adsorption of Different Functional Groups on a-C Surface Atomic Motifs

	MAE (meV)	RMSE (meV)
–H	227	313
–COOH	243	316
=O	261	338
=O/–O–	417	556
–OH	239	303

statistical distribution of functional groups and catalytic properties of surfaces in the near future.

V. CONCLUSIONS

We have conducted a comprehensive and systematic assessment of the various atomic motifs in amorphous carbon bulk and surfaces, based on a combination of DFT-based electronic structure simulations and ML algorithms. We have established a link between the geometrical features of the motifs and their reactivity toward experimentally relevant functional groups that contain hydrogen and/or oxygen. Our analysis reveals that, in addition to the standard classification into sp, sp^2, and sp^3 motifs, the sp and sp^2 motifs at a-C surfaces should be further split into two subgroups each. Our adsorption energy calculations show a strong correlation between the adsorption characteristics and motif geometry, and overall, they are in line with chemical intuition. On the basis of all the results discussed in the paper, we can derive an ordered list of structural motifs at a-C surfaces, with decreasing adsorption energies (i.e., decreasing reactivity) as follows: (1) C motif with one neighbor (cluster 1, most reactive), (2) bent sp motif (cluster 3), (3) straight sp motif (cluster 2), (4) sp^2 motif with longer bond distances (cluster 5), (5) sp^2 motif with shorter bond distances (cluster 4), and (6) sp^3 motif (cluster 6, negligible reactivity). Some of these motifs are so reactive that they will become passivated as soon as the a-C surface makes contact with air or moisture. These surfaces show significantly stronger reactivity toward –O functionalization than toward –H, –OH, and –COOH functionalizations. We expect these results, summarized in Table 1 (surface sites) and Table 2 (chemical reactivity), to be useful in establishing and understanding the surface chemistry of a-C and other types of disordered forms of carbon.

Finally, we have explored the ability of structural and electronic local atomic descriptors to be used for the prediction of adsorption energies on a-C. With these descriptors, we have optimized kernel functions and trained ML models that can reliably and accurately predict these adsorption energies at a very low computational cost. The newly introduced SOAP +LDOS kernel provides better predictions than a state-of-the-art structural-only kernel (SOAP), while requiring only slightly more computational effort. These results open the door for further optimization of combined structural and electronic kernels, toward highly accurate ML-based atomistic models. These ideas, which we have tested on adsorption energy prediction, can in turn be extended to general-purpose ML-based interatomic potentials, thus greatly increasing their range of applicability and impact on the field. This is a first crucial step on the way toward tackling more complex phenomena, such as heterogeneous catalysis and electrocatalysis.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.chemmater.8b03353.

(1) Details of the clustering algorithm used, (2) coordinates of the cluster medoids, (3) a summary of nonplanarity and nonlineary of sp^2 and sp motifs, respectively, (4) average LDOS for the different motifs, (5) results of the kernel optimization procedure, and (6) a comment on the performace of ML adsorption models for undercoordinated atoms (cluster 1) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mcaroba@gmail.com.

ORCID

Miguel A. Caro: 0000-0001-9304-4261
Volker L. Deringer: 0000-0001-6873-0278
Tomi Laurila: 0000-0002-1252-8764

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Funding from the Academy of Finland (Project 285526) and the computational resources provided for this project by CSC-IT Center for Science, Finland, are gratefully acknowledged. M.A.C. acknowledges funding from the Academy of Finland Postdoctoral Researcher program, under Grant 310574. V.L.D. acknowledges a Leverhulme Early Career Fellowship and support from the Isaac Newton Trust.

■ REFERENCES

(1) Laurila, T.; Sainio, S.; Caro, M. A. Hybrid carbon based nanomaterials for electrochemical detection of biomolecules. *Prog. Mater. Sci.* **2017**, *88*, 499–594.

(2) Nomura, K.; Takagi, A.; Kamiya, T.; Ohta, H.; Hirano, M.; Hosono, H. Amorphous oxide semiconductors for high-performance flexible thin-film transistors. *Jpn. J. Appl. Phys.* **2006**, *45*, 4303–4308.

DOI: 10.1021/acs.chemmater.8b03353
Chem. Mater. 2018, 30, 7446–7455

(3) Kamiya, T.; Hosono, H. Material characteristics and applications of transparent amorphous oxide semiconductors. *NPG Asia Mater.* **2010**, *2*, 15−22.
(4) Asefa, T.; Huang, X. Heteroatom-doped carbon materials for electrocatalysis. *Chem. - Eur. J.* **2017**, *23*, 10703−10713.
(5) Hu, S.; Richter, M. H.; Lichterman, M. F.; Beardslee, J.; Mayer, T.; Brunschwig, B. S.; Lewis, N. S. Electrical, photoelectrochemical, and photoelectron spectroscopic investigation of the interfacial transport and energetics of amorphous TiO_2/Si heterojunctions. *J. Phys. Chem. C* **2016**, *120*, 3117−3129.
(6) Mahato, N.; Banerjee, A.; Gupta, A.; Omar, S.; Balani, K. Progress in material selection for solid oxide fuel cell technology: A review. *Prog. Mater. Sci.* **2015**, *72*, 141−337.
(7) Seh, Z. W.; Kibsgaard, J.; Dickens, C. F.; Chorkendorff, I.; Nørskov, J. K.; Jaramillo, T. F. Combining theory and experiment in electrocatalysis: Insights into materials design. *Science* **2017**, *355*, eaad4998.
(8) Robertson, J. Diamond-like amorphous carbon. *Mater. Sci. Eng., R* **2002**, *37*, 129−281.
(9) Sainio, S.; Nordlund, D.; Caro, M. A.; Gandhiraman, R.; Koehne, J.; Wester, N.; Koskinen, J.; Meyyappan, M.; Laurila, T. Correlation between sp^3-to-sp^2 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.
(10) Peltola, E.; Heikkinen, J. J.; Sovanto, K.; Sainio, S.; Aarva, A.; Franssila, S.; Jokinen, V.; Laurila, T. SU-8 based pyrolytic carbon for the electrochemical detection of dopamine. *J. Mater. Chem. B* **2017**, *5*, 9033−9044.
(11) Heien, M. L. A. V.; Phillips, P. E. M.; Stuber, G. D.; Seipel, A. T.; Wightman, R. M. Overoxidation of carbon-fiber microelectrodes enhances dopamine adsorption and increases sensitivity. *Analyst* **2003**, *128*, 1413−1419.
(12) Deringer, V. L.; Caro, M. A.; Jana, R.; Aarva, A.; Elliott, S. R.; Laurila, T.; Csányi, G.; Pastewka, L. Computational surface chemistry of tetrahedral amorphous carbon by combining machine learning and density functionel theory. *Chem. Mater.* **2018**, DOI: 10.1021/acs.chemmater.8b02410.
(13) Deringer, V. L.; Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2017**, *95*, 094203.
(14) Deringer, V. L.; Csányi, G.; Proserpio, D. M. Extracting crystal chemistry from amorphous carbon structures. *ChemPhysChem* **2017**, *18*, 873−877.
(15) 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.
(16) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 184115.
(17) 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.
(18) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825−2830.
(19) 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.
(20) 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.
(21) Caro, M. A. Fork of C. Bauckhage's k-Medoids Python implementation for enhanced medoid initialization. http://github.com/mcaroba/kmedoids (accessed August 30, 2018).
(22) Bauckhage, C. Numpy/scipy Recipes for Data Science: k-Medoids Clustering. Technical Report; University of Bonn: Bonn, Germany, 2015.
(23) Caro, M. A.; Schulz, S.; O'Reilly, E. P. Origin of nonlinear piezoelectricity in III-V semiconductors: Internal strain and bond ionicity from hybrid-functional density functional theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *91*, 075203.
(24) Saff, E. B.; Kuijlaars, A. B. J. Distributing many points on a sphere. *Mathematical Intelligencer* **1997**, *19*, 5−11.
(25) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 235−242.
(26) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953−17979.
(27) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 1758−1775.
(28) Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Dułak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H. A.; Kristoffersen, H. H.; Kuisma, M.; Larsen, A. H.; Lehtovaara, L.; Ljungberg, M.; Lopez-Acevedo, O.; Moses, P. G.; Ojanen, J.; Olsen, T.; Petzold, V.; Romero, N. A.; Stausholm-Møller, J.; Strange, M.; Tritsaris, G. A.; Vanin, M.; Walter, M.; Hammer, B.; Häkkinen, H.; Madsen, G. K. H.; Nieminen, R. M.; Nørskov, J. K.; Puska, M.; Rantala, T. T.; Schiøtz, J.; Thygesen, K. S.; Jacobsen, K. W. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Phys.: Condens. Matter* **2010**, *22*, 253202.
(29) Mortensen, J. J.; et al. GPAW: DFT and beyond within the projector-augmented wave method. https://wiki.fysik.dtu.dk/gpaw/ (accessed August 30, 2018).
(30) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(31) Tkatchenko, A.; Scheffler, M. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.* **2009**, *102*, 073005.
(32) Monkhorst, H. J.; Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **1976**, *13*, 5188−5192.
(33) Xia, H.; Li, W.; Song, Y.; Yang, X.; Liu, X.; Zhao, M.; Xia, Y.; Song, C.; Wang, T.-W.; Zhu, D.; Gong, J.; Zhu, Z. Tunable magnetism in carbon-ion-implanted highly oriented pyrolytic graphite. *Adv. Mater.* **2008**, *20*, 4679−4683.
(34) Santos, E.; Schmickler, W. Electrocatalysis of hydrogen oxidation − theoretical foundations. *Angew. Chem., Int. Ed.* **2007**, *46*, 8262−8265.
(35) Newns, D. M. Self-consistent model of hydrogen chemisorption. *Phys. Rev.* **1969**, *178*, 1123−1135.
(36) Savazzi, F.; Risplendi, F.; Mallia, G.; Harrison, N. M.; Cicero, G. Unravelling some of the structure-property relationships in graphene oxide at low degree of oxidation. *J. Phys. Chem. Lett.* **2018**, *9*, 1746−1749.
(37) Sque, S. J.; Jones, R.; Briddon, P. R. Structure, electronics, and interaction of hydrogen and oxygen on diamond surfaces. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2006**, *73*, 085313.
(38) 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.
(39) Bartók, A. P.; Csányi, G. Gaussian approximation potentials: A brief tutorial introduction. *Int. J. Quantum Chem.* **2015**, *115*, 1051−1057.

DOI: 10.1021/acs.chemmater.8b03353
Chem. Mater. 2018, 30, 7446−7455