

## ARTICLE OPEN



# Molecular identification with atomic force microscopy and conditional generative adversarial networks

Jaime Carracedo-Cosme<sup>1,2</sup> and Rubén Pérez<sup>2,3</sup>✉

Frequency modulation (FM) atomic force microscopy (AFM) with metal tips functionalized with a CO molecule at the tip apex (referred as High-Resolution AFM, HR-AFM) has provided access to the internal structure of molecules with totally unprecedented resolution. We propose a model to extract the chemical information from those AFM images in order to achieve a complete identification of the imaged molecule. Our Conditional Generative Adversarial Network (CGAN) converts a stack of constant-height HR-AFM images at various tip-sample distances into a ball-and-stick depiction, where balls of different color and size represent the chemical species and sticks represent the bonds, providing complete information on the structure and chemical composition. The CGAN has been trained and tested with the QUAM-AFM data set, that contains simulated AFM images for a collection of 686000 organic molecules that include all the chemical species relevant in organic chemistry. Tests with a large set of theoretical images and few experimental examples demonstrate the accuracy and potential of our approach for molecular identification.

npj Computational Materials (2024)10:19; <https://doi.org/10.1038/s41524-023-01179-1>

## INTRODUCTION

Atomic Force Microscopy (AFM)<sup>1</sup> in combination with dynamic operation modes<sup>2,3</sup> has become one of the key tools for imaging and manipulation of materials and biological systems at the nanoscale. Operated in the frequency-modulation mode (FM) (commonly known as Non-contact AFM), AFM achieves true atomic-scale resolution<sup>2,3</sup>. The use of metal tips functionalized with a CO molecule at the tip apex, has provided access to the internal structure of molecules with totally unprecedented resolution<sup>4,5</sup>. The main contrast mechanism for AFM with inert tips like CO is Pauli repulsion<sup>4</sup>, that is due to the overlap of the electron densities of tip and sample. This repulsive force produces positive frequency shifts – changes in the oscillation frequency of the cantilever holding the tip due to the tip-sample interaction – that are observed as bright features in the constant height AFM images above atom positions and bonds, reflecting the molecular structure. Increasingly accurate AFM simulation models<sup>6–10</sup> have been developed to explain the observed image contrast. They have contributed to elucidate the role of the CO tilting<sup>7</sup>, the influence of other contributions to the tip-sample interaction, like the electrostatic force<sup>11,12</sup>, the role of the CO-metal tip charge distribution<sup>10,13</sup>, and the interplay of the short-range chemical interaction and electrostatics in bond order discrimination and the imaging of intermolecular bonds<sup>14</sup>.

High-resolution experimental (HR) AFM images, together with the ability to address individual molecules, have paved the way for the identification of natural products—like breifussin A, where the structure of some of the fragments was known but methods like nuclear magnetic resonance (NMR) failed to provide the global structure<sup>15</sup>. HR-AFM is also key in the imaging of the intermediates (including radicals) and final products generated in on-surface reactions, shedding light into the formation processes and reaction pathways<sup>16–19</sup>. The technique has also been able to resolve more than a hundred different types of molecules in asphaltenes, the solid component of crude oil<sup>20</sup>. Molecular identification in all of the previous cases was supported by

significant information about the nature of the molecules involved, as in the case of asphaltenes, where we were dealing essentially with polycyclic aromatic hydrocarbons based on C and H atoms. Achieving molecular identification in a general case is a much more challenging problem, as it is necessary to disentangle the contribution of the bonding topology, the chemical composition and the internal corrugation of the molecule to the contrast of the HR-AFM images, coping with the presence of experimental noise and tip asymmetries. In spite of the wealth of information provided by HR-AFM experiments and the advances in the interpretation of the observed contrast, the complete identification of molecular systems, i.e., the determination of the structure and composition, solely based on HR-AFM images, without any prior information, remains an open problem.

Few works have tried to tackle this problem using artificial intelligence (AI) techniques<sup>21,22</sup> to process AFM images. Deep learning (DL) is nowadays routinely used to classify, interpret, describe and analyze images<sup>23–28</sup>, providing machines with capabilities that surpass human beings<sup>29</sup>. DL ability to recognize patterns could in principle be exploited to characterize the structure of molecular systems. Kalinin and co-workers have done pioneering work<sup>30</sup> in the application of DL<sup>31</sup> and Bayesian<sup>32</sup> methods to extract information from electron microscopy and scanning probe microscopy (SPM) images. Gordon et al.<sup>33</sup> implemented a model to automate the detection of spatially correlated patterns in varied sets of AFM images of self-organized nanoparticles. However, the problem of molecular identification based on AFM images has to face two main challenges that are intrinsic to the technique: how to achieve chemical identification within the molecule at the single atom level—as the effects of both geometry and chemical composition contribute to the determination of the 3D molecular charge density, that is ultimately responsible for the AFM contrast—and how to deal with markedly non-planar, 3D structures, given that the deflection of the CO tip as we get closer to the sample can prevent the exploration of the lower lying molecular areas. Alldritt et al.<sup>21</sup>

<sup>1</sup>Quasar Science Resources S.L., Las Rozas de Madrid, Spain. <sup>2</sup>Departamento de Física Teórica de la Materia Condensada, Universidad Autónoma de Madrid, Madrid, Spain. <sup>3</sup>Condensed Matter Physics Center (IFIMAC), Universidad Autónoma de Madrid, Madrid, Spain. ✉email: [ruben.perez@uam.es](mailto:ruben.perez@uam.es)

focused on the structure challenge and developed a convolutional neural network (CNN) whose aim was to determine the molecular geometry from HR-AFM images. Working with molecules mainly based on C and H, the performance was excellent for the structure of quasi-planar molecules, even using the algorithm directly with experimental results. For 3D structures, they were able to recover information for the positions of the atoms closer to the tip, in a height range of 150 pm. However, the discrimination of other chemical species (O, Cl) and simple functional groups (OH) produced non conclusive results.

CNNs have also been used to predict accurate electrostatic fields<sup>34</sup> from HR-AFM images, while graph neural networks (GNNs) have been applied to extract molecular graphs<sup>35</sup>.

In our previous work<sup>22</sup>, we faced the chemical identification challenge and showed the feasibility of performing a very accurate automatic molecular classification with DL techniques for a set of 60 planar molecules, that include the 10 most common atomic species in organic chemistry, using their theoretically simulated AFM images. Furthermore, we proposed a variational autoencoder (VAE)<sup>36,37</sup> based method to include the characteristic features of the experimental AFM images in the training dataset, significantly increasing the accuracy of the model tested with experimental images<sup>22</sup>. However, although this approach shows the potential to recognize both the structure and composition of molecules through AFM images, it does not come close to solving the global identification problem. The classification approach can only identify molecules included in the classes predefined in the model. The output is a vector where each component provides the probability that the input belongs to the corresponding class. If we consider a small variation of one of the molecules included in the classification set, just replacing one of the atoms by another chemical species or by a functional group, the model would not be able to classify this molecule. At best, it might identify it with the parent molecule in the classification set, but it would not be able to discriminate between the two molecules. Given the rich complexity provided by organic chemistry, even using an extremely large data set—which already poses fantastic computational requirements, as the output vector would have the dimension of the number of molecules in the data set—the model would fail to classify many of the already known or possibly synthesized molecules of interest. Thus, the challenge is to build a DL model that, trained with a large but limited number of molecules, is able to generalize and identify any possible organic molecule.

In this work, we address the problem of molecular identification from a completely new perspective, using visualization techniques that map images onto images. Image translation has been widely applied for various purposes, such as image denoising<sup>38</sup>, data compression<sup>39,40</sup>, synthetic data generation<sup>41</sup> or image segmentation<sup>42</sup>. One of the most widely accepted methods in the community for these tasks is the CGAN. This enhancement of the original generative adversarial network (GAN)<sup>43</sup> has demonstrated an outstanding ability to colorize images, reconstruct objects from edge maps, and synthesize photos from labeled maps, among other tasks<sup>44</sup>. In particular, the CGAN has played a key role in problems such as the fully convolutional translation from aerial photos to maps<sup>44</sup>, that can be considered analogous to our specific goal of molecular identification through ball-and-stick molecular depictions produced from AFM images.

The architecture of a CGAN includes two neural networks: the generator and the discriminator. The generator is responsible for converting the input images into the output ones, whereas the discriminator tries to predict whether the output image is the real one (ground truth) or has been produced by the generator. The competition between these two networks forces them to improve significantly their performance during the training. For its prediction, the discriminator compares patches of the generator's input image with its output and with the real image. Thus, these

networks specialize in translating and detecting local environments of the images respectively, making the CGAN particularly suitable for molecular identification through AFM imaging, since the contrast features induced by each atom in the images depend strongly on its chemical environment and very weakly on more distant atoms.

In our CGAN implementation, the input for the generator is a stack of 10 constant-height HR-AFM images covering the range of tip-sample distances commonly used for AFM imaging, spanning a distance variation of 100 pm. To this end, we have modified the original CGAN architecture replacing the 2D convolutions in the first layers of the generator by 3D convolutions that allow processing multiple images. Our CGAN turns the stack of AFM images into a graphical representation, the ball-and-stick depiction, where balls of different color and size represent the different chemical species and sticks represent the bonds between the atoms, providing complete information on the structure and chemical composition. The CGAN has been trained and tested with the Quasar Science Resources-Autonomous University of Madrid Atomic Force Microscopy Image Dataset (QUAM-AFM)<sup>45</sup>, an open-access dataset that includes simulations of theoretical AFM images for a collection of 686,000 molecules that include all the chemical species relevant in organic chemistry. The QUAM-AFM data set have been split into training, validation and test sets with 581,000, 24,000, and 81,000 structures, respectively. Each of these sets contain a variety of molecules with different sizes, chemical compositions and bonding configurations. An extensive test with molecules randomly chosen from the large test set, that have not been shown to the CGAN during training, shows the ability of the model to identify the structure and composition of any type of organic molecule, achieving the complete generalization of the molecular identification problem. Below, we discuss the main points of our implementation and test quantitatively its performance with a large set of theoretical images and few experimental examples taken from the literature, in order to demonstrate the accuracy and high potential of this approach for molecular identification.

The problem of molecular identification from a stack of AFM images has been previously addressed by us framing it as an image captioning challenge and using multimodal networks<sup>46</sup> to solve it. Each multimodal network (M-RNN) included a CNN for image analysis and a recurrent neural network (RNN) for language processing. The first network took as input the 3D image stack and provided the attributes, the IUPAC terms corresponding to all the chemical groups present in the molecule. The second M-RNN exploited both the 3D image stack and the attributes provided by the first M-RNN to predict the IUPAC name of the molecule, that completely describes the structure and composition of the molecule. The determination of the chemical groups within the molecule had a 95% accuracy, showing that AFM images did carry significant chemical information and that the CNN is able to retrieve it. For the prediction of the complete IUPAC name, although the model outperforms most applications of RNN to language translation, the accuracy was limited to 76% using the cumulative 4-gram BLEU metric<sup>47</sup>, the standard metric for natural language processing. This performance drop is probably related to intrinsic limitations of RNNs models and to the IUPAC formulation rules, specifically designed for humans but not particular suitable for machine learning applications. The completely different approach to molecular identification that we propose here is designed to overcome this language limitation and also to provide information from images containing groups of molecules bonded by hydrogen or halogen-bond interactions or molecular fragments that cannot be described by the IUPAC formulation.

## RESULTS

### A CGAN model to identify molecules through their ball-and-stick depictions

We use a CGAN<sup>44</sup> to identify the molecules through ball-and-stick depictions. They represent each atomic species with balls of different colors and sizes centered at the position of the atoms, and define the structure through sticks, joining the balls, that represent the chemical bonds. Our proposal is based on the fact that this representation carries chemical information not only in the balls but also through the length of the sticks, since interatomic distances depend on the chemical species and the order of the bond (e.g., single, double and triple carbon–carbon (C–C) bonds have different lengths).

The model applied for the identification is based on the implementation of the CGAN proposed in ref. <sup>44</sup>. The CGAN model is composed of two networks, known as generator and discriminator. Figure 1 shows the structure and layers of each network. We define the stack of 10 AFM images at different tip-sample distances as input to the generator and the corresponding ball-and-stick depiction as output. Our proposal differs from the original implementation in the first layers of the generator: a dropout layer with a rate of 0.5 and two 3D convolutional layers (replacing the original 2D convolutional layers) to process the image stack. A dropout layer with such a high rate is important for the model to be able to generalize and make accurate predictions when dealing with experimental images.

During the training, the networks are confronted against each other in a zero-sum game consisting of two steps. First, the generator is fed with a stack of AFM images and tries to generate the ball-and-stick representation corresponding to the molecule

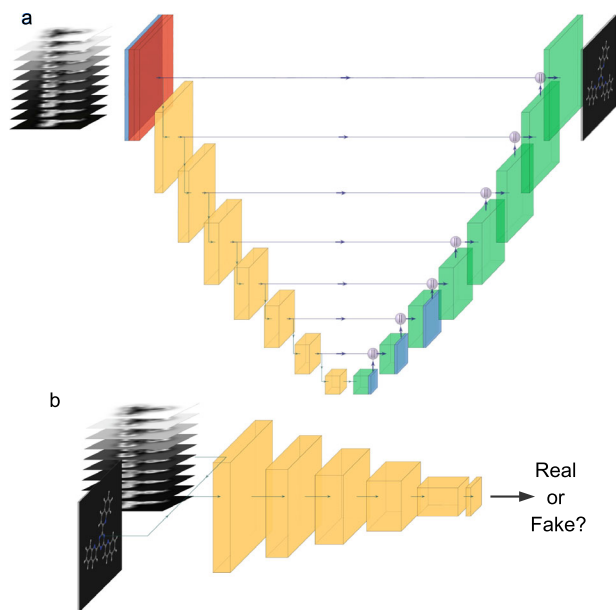
from which the input AFM images have been simulated. Second, we feed the discriminator with the AFM image stack (the same used for the generator) and also with the ball-and-stick depiction. With this data, the discriminator predicts whether the ball-and-stick depiction is the ground truth or the image generated with the generator network. In this way, we train the two networks together in an end-to-end process in which the first network learns both to fool the discriminator and to generate images as close as possible to the ball-and-stick depiction, and the discriminator learns to guess whether the second input image is real or fake. From a practical point of view, the discriminator is a network that is only useful to force the generator to improve. Therefore, once this objective has been achieved, we discard the discriminator network. The generator is in charge of generating the ball-and-stick depiction representing the atoms and bonds, providing a complete identification of the molecule.

While most of the model details are presented in the Methods section, there are two technical points that we want to highlight as they are important in order to explain the remarkable performance of our approach. The first one is related to how the discriminator makes its prediction. This is not achieved by a global assessment of the inputs but by comparing them segmented into patches of  $16 \times 16$  pixels. This local analysis based on small patches of the images makes CGAN especially powerful in AFM image analysis, as the features induced by the structure and composition on the AFM images depend strongly on the local chemical environment and smoothly on the global molecular configuration. The second one exploits the freedom to incorporate additional terms into the loss function used during the training. As suggested in the original CGAN implementation<sup>44</sup>, a distance L1 (defined as the sum of the absolute difference of the components of a vector) has been added to the loss function. This distance, an alternative to the usual Euclidean L2 norm, forces the generator not only to fool the discriminator, but also to produce outputs closer to the real ones and with as little blur as possible.

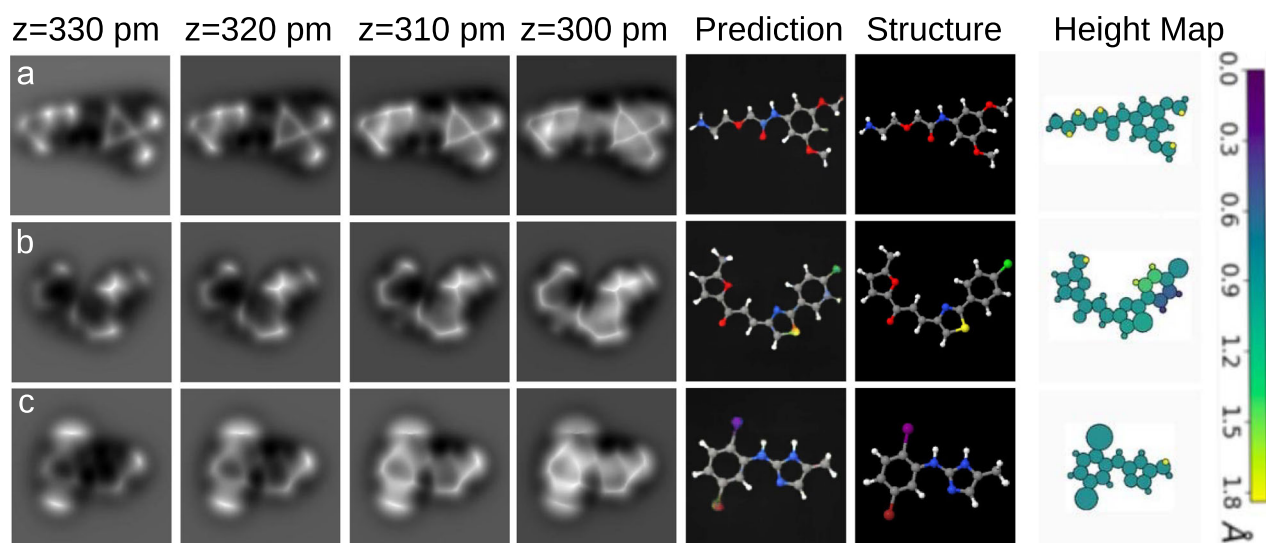
### Testing the identification with simulated AFM images

In order to evaluate the accuracy of molecular identification through AFM with the CGAN, we perform a test with 3015 structures randomly selected from the set of 81,000 molecules specifically reserved for this purpose from QUAM-AFM (see Methods). The test was not performed on the complete test set due to the fact that the evaluation was carried out by human visual comparison between the target structure and the one predicted by the model. For each of these structures, we randomized the selection of the AFM operational parameters (oscillation amplitude, lateral stiffness of the CO-metal tip bond) used in the simulations among the 24 possible combinations offered by QUAM-AFM (see Methods), resulting in 3,015 stacks of 10 tip-sample distance AFM images.

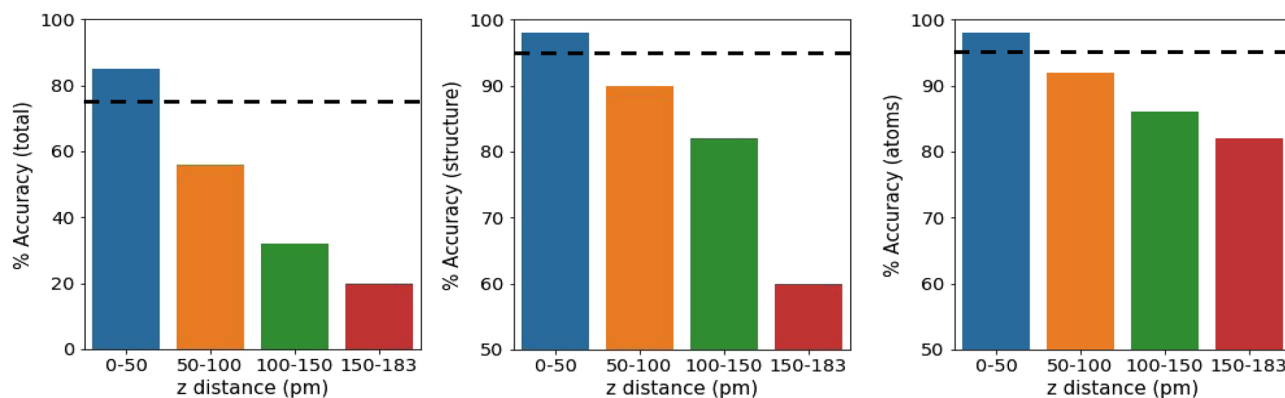
The results of the test shown in Fig. 2 demonstrate that our method works with outstanding results: theoretically simulated AFM images contain sufficient information to carry out a complete chemical and structural identification of the molecule through the prediction of its ball-and-stick depiction. The model recognizes both chemically and structurally semi-flat molecules in complex cases, including structures that a human expert would not be able to identify. Figure 2a shows the identification of 2-(2-aminoethoxy)-N-(3,5-dimethoxyphenyl)acetamide, one of these tough examples. The corresponding AFM images are characterized by strong distortions of the structure created by the strong charge accumulation around the oxygens<sup>48</sup>. These strongly electronegative atoms hide their bonds with the  $sp^3$  carbons, creating a triangular feature at the position of the ring and hiding also the presence of the nitrogen (N) atom attached to it. Nevertheless, the model is able to differentiate  $sp^3$  and  $sp^2$  carbons and identify the two amino



**Fig. 1** Our implementation of the CGAN structure. During the training, the generator model **a** and the discriminator model **b** are confronted against each other in a zero-sum game: firstly, the generator is fed with a stack of AFM images and tries to generate the ball-and-stick representation. Secondly, we feed the discriminator with the AFM image stack (the same used for the generator) and also with the ball-and-stick depiction. With this data, the discriminator has to predict whether the ball-and-stick depiction is the ground truth or the image generated with the generator network. The models include 3D convolutional layers (red boxes), dropout layers (blue), blocks of 2D convolutional layers (yellow) and with 2D transposed convolutional layers (green). For a detailed description of each block and their corresponding layers, including the activation functions, see Methods.



**Fig. 2** Examples of perfect predictions from our CGAN model. (From left to right) AFM images at different tip-sample distances, prediction performed by the CGAN, real structure and height map for **a** 2-(2-aminoethoxy)-N-(3,5-dimethoxyphenyl)acetamide, **b** 3-[2-(4-chlorophenyl)-1,3-thiazol-4-yl]-1-(5-methylfuran-2-yl)prop-2-en-1-one and **c** N-(5-bromo-2-iodophenyl)-5-methyl-1H-imidazol-2-amine.



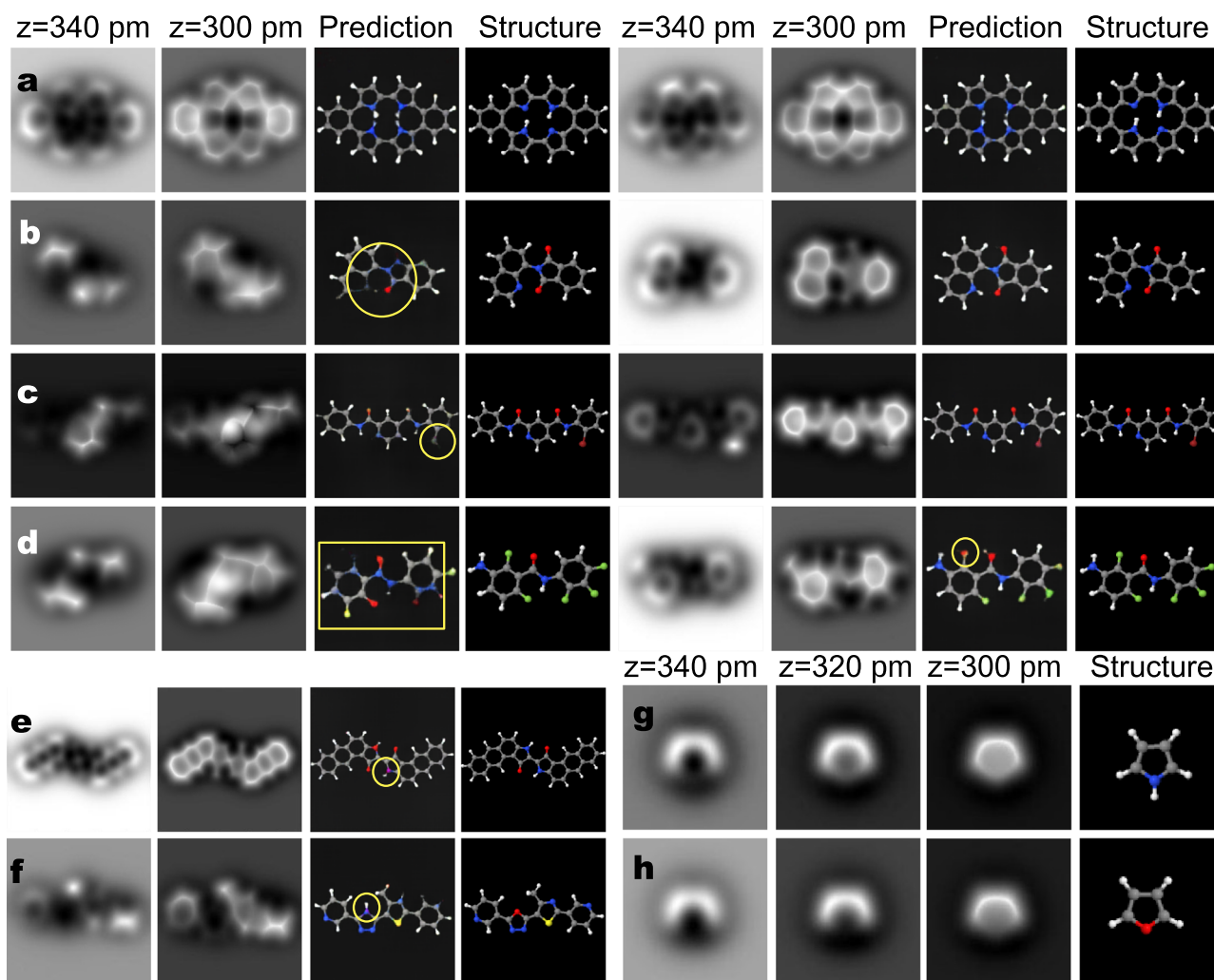
**Fig. 3** Accuracy of the model in a test where both the 3015 structures and their simulation parameters have been randomly selected. The bar charts show (from left to right) the overall accuracy (perfect structure and atom prediction), the accuracy of structure discovery, and the accuracy in revealing the atomic species. The set of structures has been divided into four subsets according to their corrugation in order to show the dependence of the model accuracy versus the height difference in the atoms of the molecule. The horizontal dashed line shows the accuracy over the complete test set. The (total) accuracy has been evaluated considering that the final result is correct only if the prediction is perfect: it shows all the bonds of the molecule, the number of vertices of each structure (chain or rings), and the proper color assigned to each atom, with the exception of the hydrogen atoms and its bonds. The structure accuracy has been calculated as the percentage of fully discovered (perfect) structures out of the total set of structures. The accuracy in the prediction of the atomic species has been evaluated as the percentage of total hits (correct predictions) over the total number of atoms in the set, without considering the hydrogen atoms. See Supplementary Table 1 for details.

groups, leading to a perfect prediction. Figure 2 (b and c) shows other remarkable achievements of the model, such as the identification of  $sp^3$  carbons, sulfur, oxygen and nitrogen atoms in different chemical environments and the accurate discrimination of three different halogen species (chlorine (Cl) in Fig. 2b and iodine (I) and bromine (Br) in Fig. 2c).

Figure 3 provides a quantitative estimate of the accuracy of our identification method using a global assessment and two specific evaluations focused on either structure or composition. The model achieves a remarkable 74% of perfect predictions, that increase to 95% (96%) when considering only structure (composition). Notice that, in the total accuracy and the structure accuracy, a prediction has been considered correct only if there is a perfect match, whereas the accuracy in the prediction of each atomic species has been assessed by considering each individual atom in the molecule as correct or incorrect and assigning to that prediction the ratio between the number of correct atoms and the total

number of atoms. This method of evaluation penalizes errors in structure discovery more than in atom determination, since, in all the predictions, most of the structure is revealed correctly, providing valuable information about the molecule, in spite of being considered as incorrect in the determination of the accuracy.

We have explored the influence of the molecular corrugation—the maximum height difference of the atoms in the molecule (excluding hydrogen (H) atoms), where the height is defined as the distance between atoms measured perpendicular to the molecular plane—in the performance of the model. The force curves associated with certain atomic species in different molecular moieties are quite similar. In fact, in some cases, these curves are almost identical except for a rigid translation, equivalent to a vertical displacement of the atoms. Thus, we could expect the model to mistake some of these atoms in a non-planar structure where they are at different heights. The test set was split into four subsets according to the maximum height



**Fig. 4** Understanding the origin of some mistakes in the CGAN predictions. **a** AFM images, predictions and structures for meso-Dibenzoporphycene (mDBPC) relaxed on a Ag(111) surface (left) and on a NaCl bilayer (right), see ref. <sup>49</sup> for structural corrugation details on each surface. AFM images, predictions and structures in the gas-phase configurations (left) and in a forced planar structure (right) for the molecules **b** 2-quinolin-8-ylisoindole-1,3-dione, **c** 4-N-(2-bromophenyl)-2-N-phenylpyridine-2,4-dicarboxamide and **d** 3-amino-2,6-difluoro-N-(2,3,4-trifluorophenyl)benzamide. AFM images, predictions and structures for **e** 2-(1-oxo-3H-naphtho[2,3-e]indol-2-ylidene)-3H-naphtho[2,3-e]indol-1-one and **f** 2-(4-methyl-2-pyridin-3-yl-1,3-thiazol-5-yl)-5-pyridin-3-yl-1,3,4-oxadiazole in a gas-phase configuration. **g, h** Comparison of AFM images of pyrrole and furan.

difference and the accuracy was evaluated independently for each subset. According to Fig. 3, both the total and the composition accuracy decrease linearly with the maximum height difference, while the structure accuracy shows this linear behavior in the range [0, 150] pm but has a stronger decay from 150 pm onwards.

Figure 4 provides some important hints on the origin of the limitations of the model revealed by the statistical analysis presented above. Starting with the role of the maximum height difference, the left panels in Fig. 4b–d show two representative AFM images, the prediction and the real structure for three molecules that have a strong internal corrugation in their gas-phase configuration. These images show that the model perfectly identifies chemically and structurally the top part of the molecules, but fails with the bottom, where the CO tip cannot retrieve enough information during constant height imaging, even at the shortest tip-sample distances, due to the CO lateral relaxation. These results explain the lower accuracy of the model for the molecules with stronger corrugation, particularly in the case of the structure accuracy, that requires a perfect identification of the whole molecular structure. At the same time, it seems to confirm

that there is a limit beyond which it is not possible to obtain information from an AFM with the current operation setups and with a single adsorption orientation of the molecule<sup>21</sup>.

We do not expect this limitation to be so crucial when dealing with the identification based on experimental images of the quasi-planar molecules (with internal corrugation below 183 pm) that we have been considering so far. In experiments, the molecules are deposited on a substrate. The final adsorption configurations are significantly flatter than the gas-phase ones, as the attractive molecule-substrate interaction compensates the steric hindrance effects responsible for the internal corrugation, even in the low reactive substrates commonly used for AFM experiments. This idea has been tested with the three molecules in Fig. 4b–d. The left panels of Fig. 4c show that in the gas phase structure, the model correctly predicts that bromine is a halogen (by bond length and ball size) but does not determine the color of the ball. A similar case is presented in Fig. 4d, where several atoms are misclassified. We have forced these three molecules to acquire a flat structure. The corresponding AFM images, the new prediction and the structure are shown on the right panels of Fig. 4b–d. The

prediction becomes perfect with respect to the structure in all of the three cases, and, composition-wise, fails only in a single atom in the case displayed in Fig. 4d.

After the analysis presented above, it is sensible to ask if the choice of training the model with the structures in QUAM-AFM, that correspond to gas-phase configurations, is the best strategy for molecular identification based on experimental images. This choice has been taken in the first place to make the simulation computationally feasible, as it is simply not possible to perform the relaxations needed to determine the adsorption configurations of all the molecules in the data set on a number of different substrates. However, our choice, more than a practical consideration, is actually guided by the fact that the AFM contrast of the different chemical species is strongly influenced by the chemical environment. Training the model with the molecular structures in QUAM-AFM, that, in general, do not correspond to the adsorbed configuration in the experiments, provides the model with the necessary information to learn the local relationships that the different chemical species may have depending on the height. Instead of learning to identify a structure in one particular configuration, the model is learning to relate atoms to their surroundings, allowing it to recognize molecules in different configurations.

Figure 4a demonstrates this idea. It shows the AFM images calculated for the stable adsorption configuration of meso-Dibenzoporphycene (mDBPC) on two different substrates: a more reactive silver Ag(111) surface and a rather inert sodium chloride (NaCl) bilayer. The final structures are quite different and neither of them is flat. This reflects in the different AFM contrast, that is in excellent agreement with the experiments in both substrates<sup>49</sup>. When the stack of images corresponding to these two configurations is shown to our model, the prediction for the structure and composition of the molecule is perfect in both cases, except for the position of the two internal hydrogen atoms that are always very difficult to determine from AFM experiments. This example with theoretical images and the experimental cases discussed below show that the training with the highly corrugated gas-phase configurations, although not enough to keep its global accuracy in the tests performed with molecules with strong corrugations, is actually an important asset of the model. These structures are making the model robust by showing how features associated with atomic species and molecular moieties evolve with the variation of height in different chemical environments. The choice of the molecular adsorption configurations on a particular substrate for training may lead the model to specialize excessively and lose the ability to generalize and identify the same molecule adsorbed on a different substrate. The gas-phase structures, combined with the choice of images generated with different AFM operational parameters and the use of an image data generator (IDG) (see Methods), introduce enough variability during the training to allow the model to identify the molecule, despite the differences introduced by the substrate. In summary, the corrugation of the gas-phase structures, rather than being a limitation, is enhancing the ability of the model to generalize and to recognize molecules in different adsorption configurations.

Beyond the subtleties in the AFM contrast created by the interplay of the chemical nature of the atoms, their chemical environment and their relative height, we have identified some misclassifications that occur with some frequency, even in rather flat configurations. Figure 4e, f shows two examples where the model swaps a N-H group in a pentagon for an oxygen (O) atom. In this case, although chemically they have different properties, the fact that the atoms are very electronegative and have a similar charge distribution reflects in the similar features they show in the AFM simulations in a perfectly planar configuration (see Fig. 4g, h). This fact makes them extremely difficult to identify in the presence of small variations in height. Another pair that is frequently mistaken for variations in height is O and fluorine (F) atoms when

connected to an aromatic ring (see Fig. 4d). This case is more surprising since, even though the two atoms are highly electronegative and of similar size, the O atom double bonded to a C atom of an aromatic ring should, at first, show some distinctive feature with respect to a C-F pair. Although the features associated to F and O atoms are similar, one would expect them to be distinguishable in a planar structure. It is not clear whether this error is due to some unknown effect on the structure or, perhaps, as they have similar sizes in the ball-and-stick representation, the model mistakes them under certain conditions.

### Molecular identification based on experimental AFM images

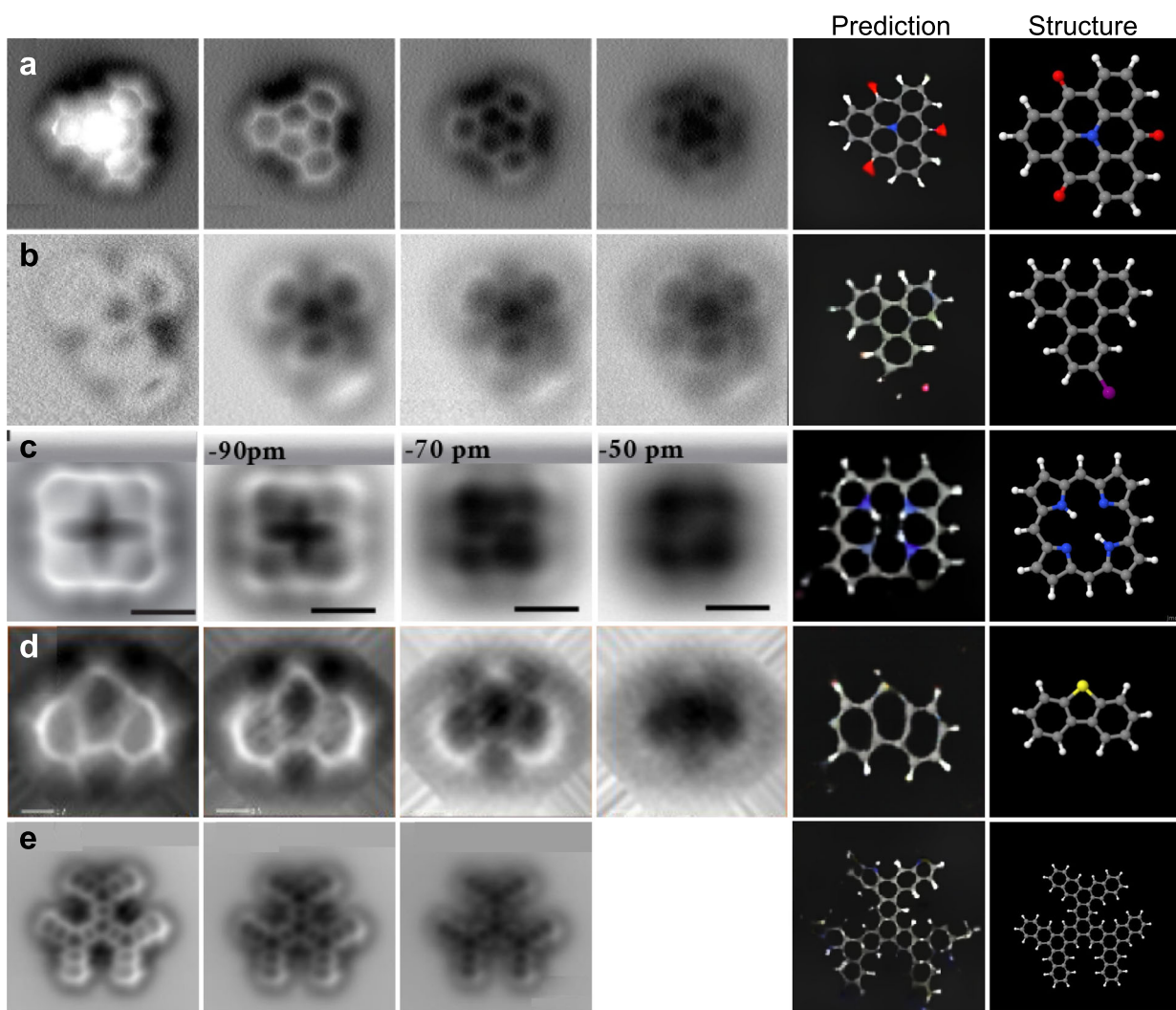
The final goal of our CGAN model is to identify molecules from their experimental AFM images. As discussed above, the range of AFM operational parameters used to simulate the images generated for each of the molecules and the use of gas-phase configurations introduce enough variability during the training to allow the model to identify the molecule, despite the differences introduced by the substrate. We have explicitly tested this point with theoretical AFM images generated for the adsorption configurations of mDBPC on two different substrates with quite different reactivity, a Ag(111) surface and a NaCl bilayer (see Fig. 4a). The theoretical AFM images faithfully reproduced the experimental results<sup>49</sup>.

Now, we want to assess the accuracy of the model with experimental results. This test is going to be limited by the scarce number of published AFM studies that include sets of images as a function of the tip height. Furthermore, most of these few studies neither provide sufficient images (10 images, taken at 10 pm intervals) nor are in the range of tip-sample distances (280–370 pm) which our analysis with simulated images have shown necessary to properly sample the variation of the tip-sample interaction and achieve complete chemical identification. Despite these drawbacks, the results presented below are really promising.

To test the performance of the model with experimental results, we have selected sets of AFM images originally published in refs. 50–55. In general, fewer than ten images corresponding to different tip-sample distances were published in these papers, so we have linearly interpolated the images two by two to extract additional images to complete the input, the stack of 10 images, required for the CGAN model. In some cases the experimental results were so limited, that it was necessary to weigh differently each image to obtain multiple results from each image pair (see Fig. 5 and Supplementary Figs. 1 and 2). We have denoised the generated 10-image stack by applying the *medianBlur* filter with size 3 from the OpenCV Python package.

It is important to stress that the interpolated images are generated for the sole purpose of completing the input dimensions required by the model, i.e., they do not provide additional information to that supplied by the original images. Therefore, the test with experimental images is really tough: We are not only increasing the complexity by using as inputs experimental images—simply cut and edited from different publications—and that, in spite of the applied filter, always carried some noise—, but we are also severely reducing the amount of information with which we feed the model.

A drawback that may hinder chemical identification by experimental AFM imaging is that the observed interaction depends on the details of the tip structure, like the attachment of the CO molecule to the metal tip. Figure 5a shows experimental AFM images, taken at constant height and acquired with a CO-terminated tip, for a 1-azahexacyclo[11.7.1.1<sup>3,19</sup>.0<sup>2,7</sup>.0<sup>9,21</sup>.0<sup>15,20</sup>]docosa-2,4,6,9(21),10,12,15,17,19-nonaene-8,14,22-trione molecule adsorbed on a Cu(111) surface<sup>50</sup>. These AFM images (and, by inheritance, also their interpolations) show an imperfect threefold symmetry. Although this asymmetry could be related to the adsorption configuration of the molecule, the discussion in ref. 50



**Fig. 5 Test of the CGAN performance with experimental AFM images.** Each panel includes the published experimental AFM images, the prediction performed with the CGAN and the structure for the corresponding molecule. **a** 1-azahexacyclo[11.7.1.1<sup>3,19</sup>.0<sup>2,7</sup>.0<sup>9,21</sup>.0<sup>15,20</sup>]docosa-2,4,6,9(21),10,12,15,17,19-nonaene-8,14,22-trione, **b** 2-iodotriphenylene, **c** 21,23-dihydroporphyrin, **d** dibenzothiophene and **e** [19]dendriphene. Experimental images in (**a**, **c**, **d**, **e**)<sup>50–54</sup> were taken in the FM operation mode at constant height, while, in **b**<sup>51</sup>, a novel Q-control AM-AFM mode was used. The color code for the balls representing the chemical species is: carbon (gray), hydrogen (white), oxygen (red), iodine (purple) and nitrogen (blue). Only 4, 6, 6, 10 and 3 experimental images were published for the molecules in panels **a–e**, respectively. Thus, when necessary, we have linearly interpolated these images in order to produce the 10-image stack used by our CGAN model as input (see text and Supplementary Figs. 1 and 2 for details). Images are reproduced from references<sup>50–53</sup> with the permission of the American Chemical Society (ACS, Copyright (2016)), AIP Publishing, American Association for the Advancement of Science (AAAS), and John Wiley and Sons. Images from reference<sup>54</sup> are reproduced under an ACS AuthorChoice License.

proves that it is really caused by the flexibility of the CO–Cu bond coupled with an asymmetric tip. Therefore, the chemical identification of this molecule has two additional complications, besides the lack of input data and the switch to experimental images: First, this structure is not part of the training set, so, in addition to testing the model with an experimental image, this is a perfect example to verify its ability to generalize. On the other hand, because in the theoretical simulations tip irregularities are not considered, the model has not been trained with images containing characteristic features induced by these asymmetrical tips in the experimental images. Despite these drawbacks, the CGAN is not only able to reveal the molecular structure but also to predict with perfect accuracy the chemical species that make up the molecule.

Besides being robust against tip asymmetries, the model seems to perform, in some case, even better in the determination of the chemical composition with experimental images than with

theoretical ones. As discussed above, one of the most common errors in the tests performed with simulated images was to mistake an O atom for an F atom in complex molecules, as they produced a similar AFM contrast. However, in the prediction of this molecule through the experimental AFM images, where the symmetry is affected by the irregularity of the tip, the model identifies the three oxygens with absolute accuracy (see Fig. 5a). It is not possible to make a general statement since the test with oxygens is limited to their presence in this particular structure, but this result seems to indicate that our CGAN is able to clearly differentiate some chemical species, like oxygens and fluorines, in experimental images.

Our CGAN model seems to work also with constant-height images taken using different AFM operation modes. Figure 5b shows the prediction performed for 2-iodotriphenylene on Ag(111) with a stack of AFM images taken using the measured oscillation amplitude in a new operation mode, Q-controlled

Amplitude Modulation Atomic Force Microscopy (AM-AFM) with CO-functionalized tips operated in constant-height mode, proposed in ref. <sup>51</sup>. The AFM images resulting from using both phase modulation in Q-control AM-AFM and frequency modulation (FM) modes on the same molecule as well as the respective predictions performed by the model are shown in Supplementary Fig. 1. As described in the Supplementary Discussion, none of these AFM images correspond to the AFM operation mode used to simulate the AFM images employed in the training of the model. This is clear for the amplitude (Fig. 5b) and phase images, but it is also the case in the FM images, as the oscillation amplitude is very different (varying from 45 to 525 pm) in each of the experimental images, while the 10-image stacks used in the training correspond to different tip-sample distances of closest approach but to the same oscillation amplitude. Moreover, the tip-height range covered by the images (64 pm) is significantly smaller than the 100 pm that we consider optimal and has been chosen so that similar contrast features were shown in the amplitude, phase and FM images. Finally, we have included in our analysis the amplitude image at the closest distance, that shows a significantly different contrast.

In spite of these severe limitations in the input, the model fed with the amplitude images fully reveals the molecular structure and the presence of the iodine atom. In the case of phase and FM images, the model gives a good description of the molecular structure but fails to provide a clear prediction about the halogen, since the color is more like the one associated to bromine than the one corresponding to iodine (See Supplementary Fig. 1). Far from considering these predictions a failure, these results indicate that our CGAN model can provide very useful information regarding the molecular identification when fed with images taken with different AFM operation modes. Nevertheless, more work is needed to reach a final conclusion about the merits and limitations of our model for this particular case, 2-iodotriphenylene on Ag(111), as shown by the analysis of another series of constant-height images taken in the frequency modulation mode for the molecule and for the products of a dehalogenation reaction locally triggered using a voltage: a triphenylene (TP) radical and the cleaved I atom<sup>55</sup> (see Supplementary Discussion). The image features at the halogen position and its evolution with tip height in Supplementary Fig. 2a are quite different from those shown in other experimental examples and from our AFM simulations, and the model predicts a methyl group instead of a halogen. In the case of the dehalogenation products (Supplementary Fig. 2b), our model captures the presence of the cleaved I atom and provides a strongly deformed structure where the dehalogenated ring is not closed, consistent with the lack of information in the AFM images due to the strong bending of the molecule towards the substrate induced by the interaction of the unsaturated C bond in that ring with the metal.

Figure 5c shows another rather successful identification, in this case, a 21,23-dihydroporphyrin molecule. The test has been carried out with interpolations from five experimental images that cover tip-sample distances varying in a range of 100 pm, although the average distance seems to be larger than the one used in the simulations of QUAM-AFM. The model is able to reveal the four pentagonal rings and the position of the nitrogens.

The predictions of the model have not been so accurate in all experimental tests. Figure 5d, e shows the test performed with AFM images of dibenzothiophene and [19]Dendriphene respectively. In the dibenzothiophene prediction, the model gets right both the number of rings and the number of vertices in each ring, which is clear in the AFM images taken at shorter tip-sample distances. However, the model is not able to rescale the central ring to show the bonds with their correct size. Furthermore, although the model manages to reveal a slight yellow color at the sulfur apex, the size of the bonds in the prediction is larger than in the target, so the prediction is not conclusive. It has to be noticed

that, despite applying a filter, we were not able to remove the experimental noise completely. Furthermore, the central ring appears, for some unknown reason, much more deformed than in the theoretically simulated images. These two features of the experimental images may account for the failure of the prediction. However, our previous work<sup>22</sup> shows that these problems with experimental images can be fixed. We proposed a strategy that significantly improves the accuracy in the classification of a small set of molecules, including dibenzothiophene, from experimental images. We implemented and trained a VAE to generate, from just three experimental images, a set of 540 images that incorporate characteristic features of the experimental cases. This subset represented a very small fraction (0.17%) of the total training set but its inclusion produced an increase in the accuracy of 0.28 (from 0.62 to 0.90) in the particular case of dibenzothiophene and an increase of 0.2 for the whole set of molecules. This strategy can be extended to our CGAN model to incorporate during the training images containing experimental features in order to improve its accuracy.

The [19]dendriphene prediction is also partly a failure. Although it reveals a large part of the structure, it does not close five of the six peripheral rings. Moreover, while in most cases, the prediction of the presence of carbon atoms is correct, the model tints some areas of the structure with bluish tones that do not allow to conclusively determine whether the chemical species is a carbon or a nitrogen. It has to be noticed that the test has been performed with only three experimental images, that is, less than a third of the information with which the model was trained. At the same time, it is also remarkable, that, even for such a complicated test and with a very limited input information, the number of vertices of each revealed ring is correct.

## DISCUSSION

In summary, our results show the potential for chemical and structural identification of molecules encoded in HR-AFM images. We propose a CGAN to generalize the accurate classification of a small set of molecules achieved in our previous work<sup>22</sup> into a general purpose tool to completely determine the structure and composition of arbitrary quasi-planar organic molecules. Our model performs a direct translation between a stack of 10 constant-height HR-AFM images and the ball-and-stick depiction of the molecule. We are only limited by the fact that the atoms composing the molecule have to be in the training dataset. Since QUAM-AFM<sup>45</sup> includes the most relevant chemical species in organic chemistry, the model prediction is practically unconstrained.

Molecular identification in both theoretical and experimental images is highly accurate with a model trained exclusively with theoretical images. The few results presented for molecular identification based on experimental HR-AFM images, in spite of the incomplete information available, are really remarkable. The ability of the model to reveal molecular structures and chemical species is truly remarkable, beyond the capabilities of a human expert in the field. Moreover, these identifications are not conditioned to a single molecular configuration, since the differences in height of the atoms in the gas-phase structures included in the training dataset provide enough information to identify patches of the image according to the chemical environment of each atom. In this way, the model has learned to decipher the distortions produced by each chemical species in relation to its surroundings regardless of the relative height difference in the molecule. Given that the presence of side functional groups and different chemical species lowers significantly the molecular symmetry, that it is further reduced during the adsorption process, we do not expect methods dealing with symmetry identification in the presence of noise<sup>56,57</sup> to be helpful in improving the performance of DL models in the identification of



molecules from experimental HR-AFM images. However, they can be certainly relevant for the scanning probe community working of the characterization of surfaces of novel materials and 2D supramolecular structures.

We attribute the high performance of the model to the consistency and robustness shown by CNNs in the analysis of images with DL, together with the patch analysis performed by the discriminator and the use of a suitable loss function, with an L1 distance, that increase the sharpness of the predictions and makes the mapping between input and output accurate. The reduced accuracy shown for structures that have a very high internal corrugation is not a critical issue when facing the identification from experimental AFM images, as real adsorbed structures tend to be flatter than the corresponding gas-phase ones. Moreover, in these high-corrugation cases, the model correctly reveals both the structure and the chemical species located on the top areas of the molecule. The presence of atoms in the lower areas is indicated with bonds that are eventually blurred due to the lack of information. Thus, more than a problem of the model, this reduced accuracy represents an intrinsic limitation of the current AFM set-ups, that may be fixed by an alternative operation mode. In this regard, a recent work<sup>58</sup> has shown how the limitations of AFM with bulky molecules can be overcome with the combination of AFM imaging with Bayesian Inference and DFT calculations in order to determine the adsorption configurations for a known molecule. Future work should explore whether a combination of this strategy with our models is able to extend the molecular identification to highly-corrugated structures.

## METHODS

### QUAM-AFM data set

DL models need large datasets to adjust the weights in each of their layers. In this work, we take advantage of QUAM-AFM<sup>45</sup>, an open-access dataset that includes simulations of theoretical AFM images, based on the latest HR-AFM modeling approaches<sup>14,45,59</sup>, for a collection of 686000 molecules that include 10 different atomic species (C, H, N, P, O, S, F, Cl, Br, I). Here we provide the main characteristics that are relevant for our study and refer the reader to the original publication<sup>45</sup> for details. QUAM-AFM focuses on quasi-planar molecules, that is, molecules which display height variations up to 183 pm along the z-axis in order to include aliphatic chains and  $sp^3$  carbon atoms (methyl groups) as possible side groups. Notice that this choice is consistent with our goal to achieve a complete molecular identification, considering that, when applying the current HR-AFM constant-height mode to 3D structures, it is only possible to recover information for the positions of the atoms closer to the tip, in a height range of 150 pm<sup>21</sup>. The corrugation histogram obtained from QUAM-AFM shows that a significant fraction (73%) of the molecules are almost flat (with corrugations smaller than 25 pm) and that 17.5% are in the 75–125 pm corrugation range, with similar ratios of 4.3% (5.2%) for the 25–75 pm (125–183 pm) ranges.

The contrast of AFM images taken in the FM mode with CO-metal tips depends on parameters, such as the cantilever oscillation amplitude or the average tip-sample distance, that can be controlled during operation, and also on the tip nature, in particular, differences in the attachment of the CO molecule to the metal tip that have been consistently observed and characterized in experiments<sup>59–61</sup>. In order to cover the widest range of variants in the AFM images, QUAM-AFM was simulated with 6 different oscillation amplitudes of the cantilever (40, 60, 80, 100, 120, 140 pm), 10 tip-sample distances (280, 290, 300, 310, 320, 330, 340, 350, 360, 370 pm), an 4 values of the elastic constant describing the tilting stiffness of the CO-metal bond (0.40, 0.60, 0.80, 1.00 N/m). These 240 combinations are applied to each of the molecular structures, resulting in a total of 165 million

gray-scale images with resolution  $256 \times 256$  pixels. QUAM-AFM also provides the ball-and-stick depictions of each molecule generated from the atomic coordinates. These depictions share the same scale used in the AFM images: if we superimpose the two images, each ball of the representation is centered on the position occupied by the atom it represents in the AFM images.

### CGAN Molecular identification model

The generator for the identification of molecules through AFM images is composed of a series of similar blocks where the main difference is the number of kernels applied in each convolution and the dimensions of each input (see Fig. 1a). The input consists of a stack of 10 greyscale AFM images (a single channel). This stack is processed in a dropout layer, with a rate of 0.5, followed by two 3D convolutional layers. The first 3D convolution includes 64 kernels, each of them has (4, 3, 3) size and is applied with a stride of (3, 1, 1) and padding. The second 3D convolution also has 64 kernels but, in this case, the kernels have size (4, 4, 4) and are applied with a stride of (4, 2, 2). The output of the second convolutional layer is resized to (128,128,64) and activated with a Leaky ReLU (LReLU) function.

From this point on, the encoder consists of seven blocks, represented by yellow boxes in Fig. 1a. Each block includes a 2D convolution followed by a batch normalization and a LReLU activation function with  $\alpha = 0.2$ . All kernels of the 2D convolution have size (4, 4) and are applied with a stride of (2, 2). The 2D convolutional layers have 128, 256, 512, 512, 512, 512, and 512 kernels, taking as reference the processing direction from the one closest to the input to the one closest to the compressed representation space. The outputs of the activations are used both to feed the next block of the encoder and to feed the decoder block of the same size. The generator decoder blocks, represented by green boxes in Fig. 1a, include the following layers: a transposed convolution, a batch normalization, a dropout layer with rate 0.2 (only in the three layers closest to the space of the compressed representation, see Fig. 1), a concatenation with the output of the corresponding encoder block, and, finally, a Rectified Linear Unit Activation Function (ReLU) activation (except for the last block, the one closest to the output, that is activated with an hyperbolic tangent function). The prediction of the images is done pixel by pixel. The model predicts the three RGB components that determine the final color of a given pixel.

The discriminator (Fig. 1b) consists of a sequence of layers, initiated by a concatenation of all input images (note that we can consider the 10 AFM images as a single image with 10 channels). It is followed by a 2D convolutional layer with 64 kernels of size (4, 4) and stride of (2, 2) activated with LReLU. Then, it has four blocks consisting of a 2D convolutional layer, a batch normalization and a LReLU activation ( $\alpha = 0.2$ ). The convolutions have 128, 256, 512 and 512 kernels with size (4, 4) and stride (2, 2) respectively. The last layer is a 2D convolution with a single kernel of size (4, 4) which is activated with the sigmoid function.

### CGAN training

The 686,000 structures in QUAM-AFM have been split into training, validation and test sets with 581,000, 24,000, and 81,000 structures respectively. The test set is chosen to be particularly large for two reasons. Firstly, to perform a quantitative analysis with randomly chosen structures in order to avoid an statistical fluke. Second, it is desirable to have sufficient variety of structures to be able to show examples that reflect the most salient strengths and weaknesses of the model.

During training, we randomly choose one of the combinations of AFM simulation parameters available in QUAM-AFM for each input stack. This variability in the input data makes sure that the parameters with which the AFM experiment has been carried out do not play a decisive role in the success of the

identification, prevents overfitting, and provides the model with the ability to generalize. This variability is further enhanced with the application of an IDG to the training set. This technique, commonly used in DL, applies different deformations (zoom, rotations, shifts, flips and shear) to the input images. Let's recall that the ball-and-stick depictions included in QUAM-AFM share the same scale as the AFM images. Thus the IDG has to be applied to both the input AFM images and the ball-and-stick depiction during the training: i.e., if we rotate the input AFM images, then, the corresponding ball-and-stick depiction must be rotated with the same angle. Otherwise the atomic positions of the ball-and-stick representation would not match the corresponding atomic positions of the AFM images, and the CGAN would not be able to learn a local translation (from the pixel environment) between the shape and intensity of the AFM image and the type of atom that caused it. This applies to all the operations in the IDG except for the shear, that is not applied to the output ball-and-stick depiction. This is motivated by the fact that shear represents a deformation that may appear in the experiments due to noise or tip asymmetries but it should not be present in the prediction.

We have found that the selection of appropriate deformation parameters for the IDG applied to the training set during the fitting considerably increases the accuracy of the model in the test carried out with experimental images<sup>22</sup>. An particular example of the application of the IDG and information on the range values used for the different operations can be found in Supplementary Methods (Supplementary Fig. 3).

Regarding the loss functions, the generator of the CGAN was compiled with mean absolute error (MAE) (using the parameter  $\lambda = 100$  defined by Isola et al.<sup>44</sup>), while the binary cross entropy was used for the discriminator. The model was minimized by applying batches of 32 inputs with the Adaptive Moment Estimator (Adam) optimizer, where the learning rate and first moment parameters were set to  $2 \times 10^{-4}$  and 0.5, respectively. The training of the model was carried out during six epochs (109,000 iterations), displaying 300 predictions of the validation set to estimate the optimal training point every 10,000 iterations.

## DATA AVAILABILITY

The data sets used for the training and testing of the models are part of the QUAM-AFM data set. QUAM-AFM requires 713 Gb storage and is freely available at our university institutional repository using the following link: <https://doi.org/10.21950/UTGMZ7>.

## CODE AVAILABILITY

The detailed information provided about the CGAN (layers, blocks, kernels and activation functions) and its training is enough to recreate our model using standard open-access libraries like TensorFlow and PyTorch.

Received: 12 March 2023; Accepted: 29 November 2023;

Published online: 20 January 2024

## REFERENCES

1. Binnig, G., Quate, C. F. & Gerber, C. Atomic force microscope. *Phys. Rev. Lett.* **56**, 930–933 (1986).
2. García, R. & Pérez, R. Dynamic atomic force microscopy methods. *Surf. Sci. Rep.* **47**, 197–301 (2002).
3. Giessibl, F. J. Advances in atomic force microscopy. *Rev. Mod. Phys.* **75**, 949–983 (2003).
4. Gross, L., Mohn, F., Moll, N., Liljeroth, P. & Meyer, G. The chemical structure of a molecule resolved by atomic force microscopy. *Science* **325**, 1110–1114 (2009).
5. Pavliček, N. & Gross, L. Generation, manipulation and characterization of molecules by atomic force microscopy. *Nat. Rev. Chem.* **1**, 0005 (2017).
6. Moll, N., Gross, L., Mohn, F., Curioni, A. & Meyer, G. A simple model of molecular imaging with noncontact atomic force microscopy. *N. J. Phys.* **14**, 83023 (2012).
7. Hapala, P. et al. Mechanism of high-resolution STM/AFM imaging with functionalized tips. *Phys. Rev. B* **90**, 085421 (2014).
8. Guo, C.-S., Van Hove, M. A., Ren, X. & Zhao, Y. High-resolution model for non-contact atomic force microscopy with a flexible molecule on the tip apex. *J. Phys. Chem. C* **119**, 1483–1488 (2015).
9. Sakai, Y., Lee, A. J. & Chelikowsky, J. R. First-principles atomic force microscopy image simulations with density embedding theory. *Nano Lett.* **16**, 3242–3246 (2016).
10. Ellner, M. et al. The electric field of CO tips and its relevance for atomic force microscopy. *Nano Lett.* **16**, 1974–1980 (2016).
11. Van Der Lit, J., Di Cicco, F., Hapala, P., Jelinek, P. & Swart, I. Submolecular resolution imaging of molecules by atomic force microscopy: the influence of the electrostatic force. *Phys. Rev. Lett.* **116**, 096102 (2016).
12. Hapala, P. et al. Mapping the electrostatic force field of single molecules from high-resolution scanning probe images. *Nat. Commun.* **7**, 11560 (2016).
13. Ellner, M., Pou, P. & Pérez, R. Atomic force microscopy contrast with CO functionalized tips in hydrogen-bonded molecular layers: does the real tip charge distribution play a role? *Phys. Rev. B* **96**, 075418 (2017).
14. Ellner, M., Pou, P. & Pérez, R. Molecular identification, bond order discrimination, and apparent intermolecular features in atomic force microscopy studied with a charge density based method. *ACS Nano* **13**, 786–795 (2019).
15. Hanssen, K. Ø. et al. A combined atomic force microscopy and computational approach for the structural elucidation of Breitfussin A and B: highly modified halogenated dipeptides from *Thuaria breitfussi*. *Angew. Chem. Int. Ed.* **51**, 12238–12241 (2012).
16. de Oteyza, D. G. et al. Direct imaging of covalent bond structure in single-molecule chemical reactions. *Science* **340**, 1434–1437 (2013).
17. Kawai, S. et al. Thermal control of sequential on-surface transformation of a hydrocarbon molecule on a copper surface. *Nat. Commun.* **7**, 12711 (2016).
18. Kawai, S. et al. Competing annulene and radialene structures in a single anti-aromatic molecule studied by high-resolution atomic force microscopy. *ACS Nano* **11**, 8122–8130 (2017).
19. Schulz, F. et al. Precursor geometry determines the growth mechanism in graphene nanoribbons. *J. Phys. Chem. C* **121**, 2896–2904 (2017).
20. Schuler, B., Meyer, G., Peña, D., Mullins, O. C. & Gross, L. Unraveling the molecular structures of asphaltenes by atomic force microscopy. *J. Am. Chem. Soc.* **137**, 9870–9876 (2015).
21. Alldritt, B. et al. Automated structure discovery in atomic force microscopy. *Sci. Adv.* **6**, eaay6913 (2020).
22. Carracedo-Cosme, J., Romero-Muñoz, C. & Pérez, R. A deep learning approach for molecular classification based on afm images. *Nanomaterials* **11**, 1658 (2021).
23. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2017).
24. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Bengio, Y. & LeCun, Y. (eds.) *3rd Int. Conf. Learn. Rep.* (ICLR, Wall St, La Jolla, CA, USA, 2015).
25. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. Comput. Vision Pattern Recognit. (CVPR)*, 770–778 (IEEE Computer Society Press, Piscataway, NJ, USA, 2016).
26. Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. Inception-v4, inception-resnet and the impact of residual connections on learning. In *31rd Proc. AAAI Conf. on Artificial Intelligence*, 4278–4284 (AAAI Press, Palo Alto, CA, USA, 2017).
27. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In *Proc. Comput. Vision Pattern Recognit. (CVPR)*, 1800–1807 (IEEE Computer Society Press, Piscataway, NJ, USA, 2017).
28. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proc. Comput. Vision Pattern Recognit. (CVPR)*, 4510–4520 (IEEE Computer Society Press, Piscataway, NJ, USA, 2018).
29. He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Int. Conf. Comput. Vision (ICCV)*, 1026–1034 (IEEE Computer Society Press, Piscataway, NJ, USA, 2015).
30. Ovchinnikov, O. S., Jesse, S., Bintacchit, P., Trolier-McKinstry, S. & Kalinin, S. V. Disorder identification in hysteresis data: recognition analysis of the random-bond-random-field ising model. *Phys. Rev. Lett.* **103**, 157203 (2009).
31. Kalinin, S. V. et al. Deep learning for electron and scanning probe microscopy: from materials design to atomic fabrication. *MRS Bull.* **47**, 931–939 (2022).
32. Ziatdinov, M., Liu, Y., Kelley, K., Vasudevan, R. & Kalinin, S. V. Bayesian active learning for scanning probe microscopy: from gaussian processes to hypothesis learning. *ACS Nano* **16**, 13492–13512 (2022).
33. Gordon, O. M., Hodgkinson, J. E., Farley, S. M., Hunsicker, E. L. & Moriarty, P. J. Automated searching and identification of self-organized nanostructures. *Nano Lett.* **20**, 7688–7693 (2020).
34. Oinonen, N. et al. Electrostatic discovery atomic force microscopy. *ACS Nano* **16**, 89–97 (2022).

35. Oinonen, N., Kurki, L., Ilin, A. & Foster, A. S. Molecule graph reconstruction from atomic force microscope images with machine learning. *MRS Bull.* **47**, 1–11 (2022).
36. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. In Bengio, Y. & LeCun, Y. (eds.) *2nd Int. Conf. Learn. Rep.* (ICLR, Wall St, La Jolla, CA, USA, 2014).
37. Dilkthanakul, N. et al. Deep unsupervised clustering with gaussian mixture variational autoencoders. *arXiv preprint arXiv:1611.02648* (2016).
38. Zhang, K., Zuo, W., Chen, Y., Meng, D. & Zhang, L. Beyond a gaussian denoiser: residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **26**, 3142–3155 (2017).
39. Rippel, O. & Bourdev, L. Real-time adaptive image compression. In *International Conference on Machine Learning*, 2922–2930 (PMLR, 2017).
40. Mentzer, F., Agustsson, E., Tschannen, M., Timofte, R. & Van Gool, L. Conditional probability models for deep image compression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4394–4402 (2018).
41. Oord, A. v. d. et al. Conditional image generation with pixelcnn decoders. *arXiv preprint arXiv:1606.05328* (2016).
42. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. & Maier-Hein, K. H. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021).
43. Goodfellow, I. et al. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **27** (2014).
44. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proc. Comput. Vision Pattern Recognit. (CVPR)*, 5967–5976 (IEEE Computer Society Press, Piscataway, NJ, USA, 2017).
45. Carracedo-Cosme, J., Romero-Muñiz, C., Pou, P. & Pérez, R. Quam-afm: a free database for molecular identification by atomic force microscopy. *J. Chem. Inf. Model.* **62**, 1214–1223 (2022).
46. Carracedo-Cosme, J., Romero-Muñiz, C., Pou, P. & Pérez, R. Molecular identification from afm images using the iupac nomenclature and attribute multimodal recurrent neural networks. *ACS Appl. Mater. Interfaces* **15**, 22692–22704 (2023).
47. Papineni, K., Roukos, S., Ward, T. & Zhu, W.-J. Bleu: A method for automatic evaluation of machine translation. In *40th Proc. Annu. Meet. ACL*, 311–318 (Association for Computational Linguistics, Philadelphia, Pennsylvania, 2002).
48. Zahl, P. et al. Hydrogen bonded trimesic acid networks on cu(111) reveal how basic chemical properties are imprinted in hr-afm images. *Nanoscale* **13**, 18473–18482 (2021).
49. Shimizu, T. K. et al. Effect of molecule–substrate interactions on the adsorption of meso-dibenzoporphyrane tautomers studied by scanning probe microscopy and first-principles calculations. *J. Phys. Chem. C* **124**, 26759–26768 (2020).
50. van der Heijden, N. J. et al. Characteristic contrast in  $\Delta f_{min}$  maps of organic molecules using atomic force microscopy. *ACS Nano* **10**, 8517–8525 (2016).
51. Martin-Jimenez, D. et al. Bond-level imaging of organic molecules using q-controlled amplitude modulation atomic force microscopy. *Appl. Phys. Lett.* **117**, 131601 (2020).
52. Fatayer, S. et al. Molecular structure elucidation with charge-state control. *Science* **365**, 142–145 (2019).
53. Vilas-Varela, M. et al. [19] dendriphene: a 19-ring dendritic nanographene. *Chem. A Eur. J.* **24**, 17697–17700 (2018).
54. Zahl, P. & Zhang, Y. Guide for atomic force microscopy image analysis to discriminate heteroatoms in aromatic molecules. *Energy Fuels* **33**, 4775–4780 (2019).
55. Martin-Jimenez, D. et al. Bond-level imaging of the 3d conformation of adsorbed organic molecules using atomic force microscopy with simultaneous tunneling feedback. *Phys. Rev. Lett.* **122**, 196101 (2019).
56. Kanatani, K. Comments on “symmetry as a continuous feature”. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 246–247 (1997).
57. Moeck, P. Objective crystallographic symmetry classifications of a noisy crystal pattern with strong Fedorov-type pseudosymmetries and its optimal image-quality enhancement. *Acta Crystallogr. Sect. A* **78**, 172–199 (2022).
58. Järvi, J. et al. Integrating bayesian inference with scanning probe experiments for robust identification of surface adsorbate configurations. *Adv. Funct. Mater.* **31**, 2010853 (2021).
59. Liebig, A., Hapala, P., Weymouth, A. J. & Giessibl, F. J. Quantifying the evolution of atomic interaction of a complex surface with a functionalized atomic force microscopy tip. *Sci. Rep.* **10**, 14104–14116 (2020).
60. Weymouth, A. J., Hofmann, T. & Giessibl, F. J. Quantifying molecular stiffness and interaction with lateral force microscopy. *Science* **343**, 1120–1122 (2014).
61. Neu, M. et al. Image correction for atomic force microscopy images with functionalized tips. *Phys. Rev. B* **89**, 205407 (2014).

## ACKNOWLEDGEMENTS

We would like to acknowledge support from the Comunidad de Madrid Industrial Doctorate Programme 2017 under reference number IND2017/IND-7793 and from Quasar Science Resources S.L. R.P. acknowledges support from the Spanish Ministry of Science and Innovation, through projects PID2020-115864RB-I00, TED2021-132219A-I00, and the “María de Maeztu” Programme for Units of Excellence in R&D (CEX2018-000805-M). Computer time provided by the Red Española de Supercomputación (RES) at the Finisterrae II Supercomputer is also acknowledged.

## AUTHOR CONTRIBUTIONS

J.C.-C. and R.P. conceived the project. J.C.-C. wrote the models and performed the training and testing under the supervision of R.P. All authors discussed the results and contributed to the writing of the manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41524-023-01179-1>.

**Correspondence** and requests for materials should be addressed to Rubén. Pérez.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024