Integrating Bayesian Inference with Scanning Probe Experiments for Robust Identification of Surface Adsorbate Configurations

Jari Järvi,[1,*] Benjamin Alldritt,[1] Ondřej Krejčí,[1] Milica Todorović,[1] Peter Liljeroth,[1] and Patrick Rinke[1]

[1] *Department of Applied Physics, Aalto University, P.O. Box 11100, 00076 Aalto, Espoo, Finland*

Controlling the properties of organic/inorganic materials requires detailed knowledge of their molecular adsorption geometries. This is often unattainable, even with current state-of-the-art tools. Visualizing the structure of complex non-planar adsorbates with atomic force microscopy (AFM) is challenging, and identifying it computationally is intractable with conventional structure search. In a fresh approach, we propose to integrate cross-disciplinary tools for a robust and automated identification of 3D adsorbate configurations. We employ Bayesian optimization with first-principles simulations for accurate and unbiased structure inference of multiple adsorbates. The corresponding AFM simulations then allow us to fingerprint adsorbate structures appearing in AFM experimental images. In the instance of bulky (1S)-camphor adsorbed on the Cu(111) surface, we found three matching AFM image contrasts, which allowed us to correlate experimental image features to distinct cases of molecular adsorption.

I. INTRODUCTION

The adsorption geometry of molecular adsorbates is a key parameter controlling many on-surface properties, such as diffusion and, more generally, the mechanism and yield of heterogeneous chemical reactions [1, 2]. In the field of heterogeneous catalysis, powerful electron microscopy -based techniques are now capable of resolving the structure of the catalyst surface on the atomic scale [3, 4]. However, these methods still cannot determine the adsorption configuration of the reactant on the active site. In general, visualizing non-planar adsorption structures on the single-molecule level remains a challenging task.

The current state-of-the-art in visualizing nanostructures in atomic resolution is scanning probe microscopy. Atomic resolution can be achieved with non-contact atomic force microscopy (AFM) with functionalized carbon monoxide (CO) tips [5, 6]. CO-AFM excels in structure analysis of planar organic molecules in real space, allowing for direct identification of molecular structures [6] and conformations [7]. Existing work has primarily focused on geometrically flat (planar) species. Only a few 3-dimensional (i.e. non-planar) molecules with limited conformations have been investigated [7–17], as the interpretation of different 3D adsorbate conformations remains a considerable challenge.

First principles calculations, e.g. density-functional theory (DFT) [18, 19], are a powerful tool for simulating and identifying adsorption structures. DFT provides an accurate quantum mechanical description of important adsorbate-surface interactions, but exhaustive structure search is needed to determine all the different adsorbate structures. AFM images of 2-dimensional adsorbates can inform the structure search about the molecular registry and orientation at the substrate [15, 20, 21]. In contrast, images of complex non-planar molecules are often not conclusive enough, and estimating the structures using chemical intuition is difficult. Here, we propose a computationally efficient method to determine the structure of 3-dimensional organic adsorbates using Bayesian inference with chemical building blocks and AFM simulations.

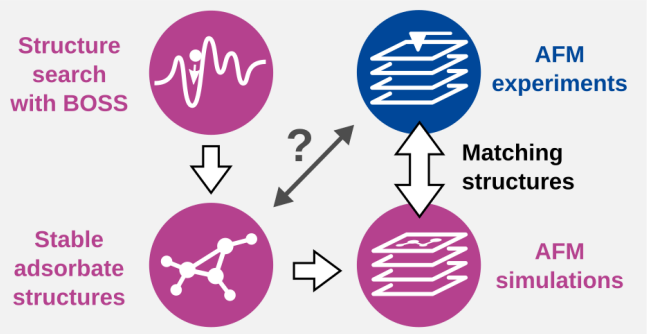

FIG. 1. **Concepts and workflow of the proposed methodology.** Identifying the structure of 3D adsorbates is often difficult from AFM experimental images (black arrow). In our combined approach, we first perform global structure search with the Bayesian Optimization Structure Search (BOSS) method and density-functional theory (DFT) to identify the stable model structures. We then simulate atomic force microscopy (AFM) images for the identified structures. We analyze the features in the simulated images and compare them to the corresponding features in experimental AFM images to detect matching configurations (white arrows).

Stable adsorbate structures can be objectively identified as the local minima of the adsorption energy landscape (AEL). Thorough sampling of high-dimensional AELs with conventional methods [22, 23] requires excessively many energy calculations, constraining us to fast force field methods which do not have the required accuracy to describe molecular adsorption. To overcome these limitations, novel Bayesian inference methods have recently been employed [24, 25]. Gaussian process regression [26] is a particularly promising technique capable of constructing a surrogate model of the AEL with a

* jari.jarvi@aalto.fi

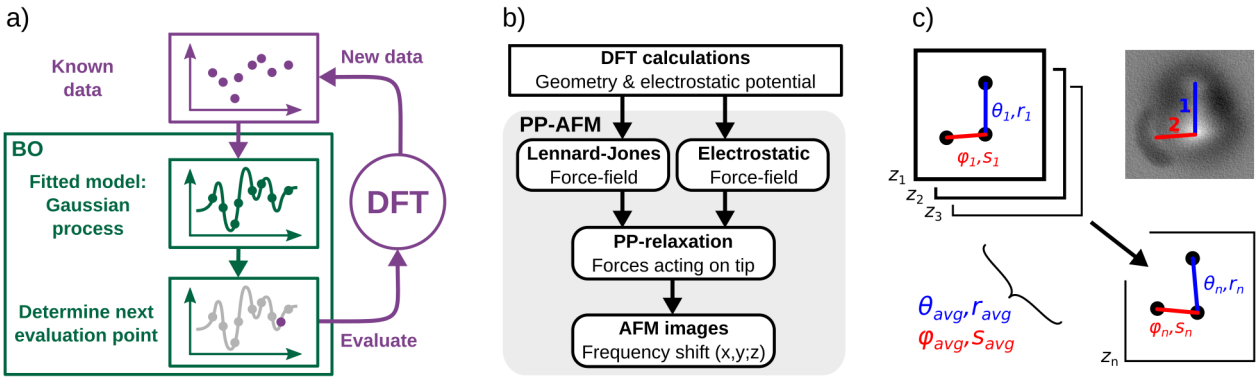

FIG. 2. **Methodology for our integrated approach.** a) Basic principle of the BOSS method, in which Bayesian optimization (BO) is applied iteratively with DFT to build a surrogate model of the AEL. In BO, the known data is first fitted in a Gaussian process, after which the next evaluation point is determined using an acquisition function. The new point is evaluated with DFT and the process is repeated with the new data included. b) Workflow of the Probe Particle (PP)-AFM simulation method. The geometry and electrostatic potential of the structure from DFT are used to compute molecular mechanic force-fields. The PP, which mimics the flexible tip-apex, relaxes in this force-field. The final force acting on the last metallic (fixed) atom of the tip is used to calculate the frequency shift Δf. c) Experiment-simulation image matching, in which AFM images are analyzed via orientations (θ, φ) and lengths (r, s) of the observed features 1 and 2. The analysis is performed on a stack of n images, obtained at different heights of the CO tip. The orientation and length of each feature is calculated as an average of the measured values in the image stack.

modest number of energy points. When combined with active learning in Bayesian optimization [27], it can be used to accelerate the construction of the AEL model via strategic sampling. The complete AEL then allows us to identify all the stable structures and estimate their mobility via the associated energy barriers. In this study, we rely on the recently developed Bayesian Optimization Structure Search (BOSS) method [28–31] to model the surrogate AEL.

Our objective here is to construct and test new methodology for automated and robust search of adsorption geometries for bulky 3D molecules. We integrate tools from different research fields to identify adsorbate structures without any requirement of previous knowledge about the studied material. Our work flow features i) global structure search with BOSS and DFT, ii) AFM image simulation with the Probe Particle (PP)-AFM model [32–34], and iii) AFM experiments (FIG. 1). Several experimental structures could be identified solely based on a single model of the AEL. We demonstrate the success and efficiency of this approach by identifying the stable adsorbate structures of (1S)-camphor ($C_{10}H_{16}O$) on the Cu(111) surface.

Previous AFM experiments [35] have shown that (1S)-camphor (a typical bulky molecule) adsorbs to Cu(111) in different stable configurations. The adsorption structure, in particular the orientation of the molecule, was difficult to interpret from the AFM images. In this work, we first use BOSS (FIG. 2a) to identify all the stable molecular adsorbate structures and their energy barriers. We select the most promising structures and generate simulated AFM images (FIG. 2b) for them. By correlating the features in experimental and simulated AFM images (FIG. 2c), we detect matches to identify several adsorbate structures observed in experiments.

II. RESULTS

A. Identifying stable adsorbate configurations

In a preparatory study [36], we applied BOSS with DFT to identify the stable adsorbate structures of (1S)-camphor on the Cu(111) surface. With BOSS, we constructed a surrogate model of the 6D AEL as a function of the orientation (3D) and location (3D) of the molecule. In the minima of the AEL, we extracted the stable adsorbate structures and their energy barriers of molecular rotation and diffusion. We identified 8 stable structures with varying molecular orientations, adsorption sites, and energy barriers.

Here, we discuss the suitability of the identified adsorbates to be observed in AFM experiments. The structures are classified in two categories, A and B, based on their adsorption properties. Class A structures, in which (1S)-camphor chemisorbs to Cu(111) via oxygen (O), are the most stable and have the highest energy barriers of molecular rotation and diffusion. In class B, (1S)-camphor physisorbs to Cu(111) via hydrocarbon interactions. These structures are less stable and the energy barriers of molecular mobility are low. In this study, we select the adsorbate structures that are the most likely to be observed with AFM due to their high stability and low mobility.

In AFM imaging, as the tip approaches the molecule, it can prompt rotation or diffusion of the molecule on the surface. Therefore, high energy barriers of molecu-

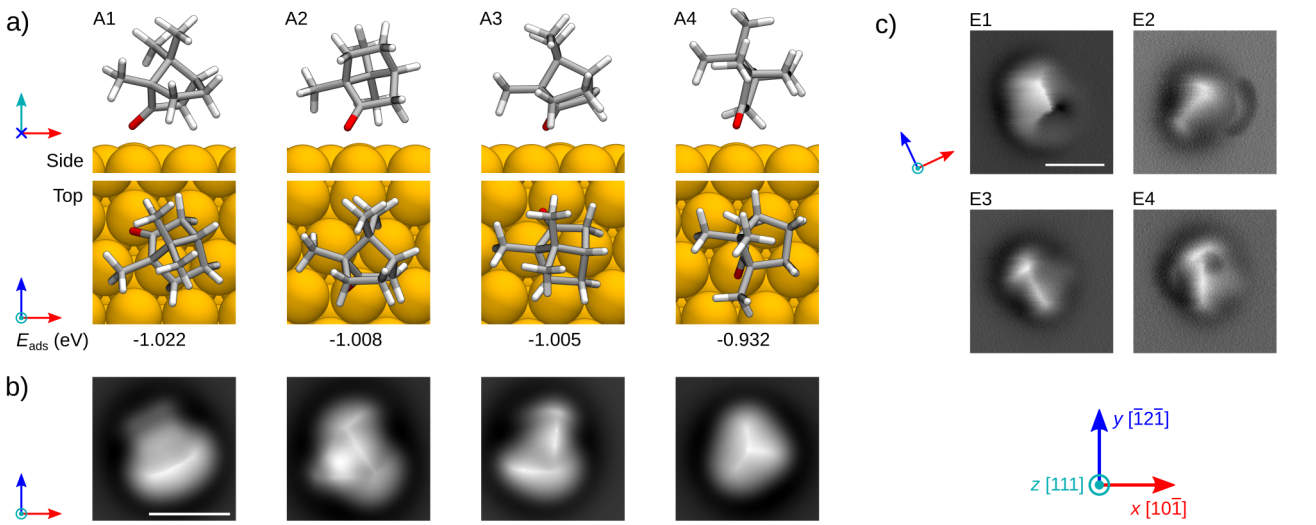

FIG. 3. **Summary of key results for stable adsorbate structures of (1S)-camphor on Cu(111).** a) Stable model structures A1–A4 and their adsorption energies (E_{ads}), predicted by BOSS and relaxed with DFT. The top views of the structures are showing area of 6.85×6.85 Å. b) Simulated AFM images of the model structures. Here is shown a single image from the image stack, taken at height 5.6 Å above the highest atom of the calculated structure. c) Constant-height AFM images, showing 4 different adsorbate structures. Coordinate axes indicate the Cu(111) lattice orientation. Scale bars in b) and c) are 5 Å.

lar motion are essential to image a static molecule with AFM. With this prerequisite, we conclude that structures A1–A4 (FIG. 3a) in our previous study [36] are the most stable and least mobile, and thus the most likely adsorbates to be observed in experiments. In these structures, the average adsorption energy (E_{ads}) of (1S)-camphor is 0.992 eV (TAB. I) and the average energy barriers of molecular rotation (E_R) and diffusion (E_D) are 0.227 and 0.029 eV, respectively. Next, we employ these structures to generate simulated AFM images, which represent the most stable adsorbates of (1S)-camphor on Cu(111).

B. Simulating AFM images

With the identified stable adsorbates, we produce simulated AFM images (FIG. 3b) for direct comparison of the structures with AFM experiments. We simulate CO-AFM with the PP-AFM method (FIG. 2b) [32–34] using different heights of the CO tip from the surface. For each structure, we obtain a stack of images, from which we then extract distinct features at different heights (FIG. 2c). The corresponding features in a stack of experimental AFM images allow us to compare simulated structures with experiments and detect matching structures.

In structures A1–A4, the simulated AFM images show varying features depending on the height of the CO tip from the molecule. The image features of a 3D molecule, such as (1S)-camphor, originate predominantly from the topmost atoms of the molecule. In the highest images of the stack, we typically observe a single point-like signal that originates from the topmost atom of the molecule. As the tip approaches the molecule, new features emerge, which manifest as additional points or a change in the shape of the previously observed features, for example their elongation. The observed features in the image stack provide us distinct fingerprints for image comparison between simulated and experimental AFM and facilitate detecting the matching structures.

For each structure, we simulated images in the height range [5.3, 6.5] Å, measured as the distance between the CO tip and the highest atom of (1S)-camphor. In the upper limit of this range, the image features begin to appear and they remain recognizable to the lower limit. The images were produced at height steps of 0.1 Å, which provides a discernible difference between images at each step.

C. Experimental AFM images

With CO-AFM, we imaged 14 adsorbed (1S)-camphor molecules on a Cu(111) surface. Approximately half of the imaged structures feature a mobile adsorbate, in which the orientation of (1S)-camphor changes during the measurement. In this study, we disregard the mobile adsorbates and focus on 4 imaged static structures (FIG. 3c), which feature distinct signals that originate from the topmost atoms of (1S)-camphor.

For comparison with simulated AFM images, we analyze the features of each structure using a stack of images (8 to 11 images per structure), measured with varying height of the CO tip from the surface. The orientation of

TABLE I. **Comparison of model structures and AFM image features.** Adsorption energy (E_{ads}) of (1S)-camphor on Cu(111), energy barriers of molecular rotation (E_{R}) and diffusion(E_{D}), distance (d_{DFT}) and orientation (θ_{DFT}) between the two topmost atoms of (1S)-camphor on Cu(111) in the model structures predicted by BOSS. Average length ($\overline{d}$) and orientation ($\overline{\theta}$) of the main features in the stack of simulated (sim) and experimental (exp) AFM images. The average lengths and orientations of all image features and their standard deviations are provided in the SM.

	BOSS/DFT					simulated AFM			experimental AFM	
	E_{ads} (eV)	E_{R} (eV)	E_{D} (eV)	d_{DFT} (Å)	θ_{DFT} (°)	$\overline{d}_{\mathrm{sim}}$ (Å)	$\overline{\theta}_{\mathrm{sim}}$ (°)		$\overline{d}_{\mathrm{exp}}$ (Å)	$\overline{\theta}_{\mathrm{exp}}$ (°)
A1	-1.022	0.232	0.045	1.77	25.2	2.76	27.9	E1	2.74	7.8
A2	-1.008	0.216	0.034	2.23	3.9	3.82	3.0	E2	3.55	27.4
A3	-1.005	0.183	0.008	2.54	24.3	4.04	29.1	E3	3.77	38.1
A4	-0.932	0.278	0.027	1.77	19.6	3.16	23.7	E4	4.00	13.8

(1S)-camphor is given with respect to the nearest crystallographic axis of the Cu(111) surface lattice in the clockwise direction, in the range $[0, 60]^\circ$. For the angle determination, we note that the Cu lattice is rotated by $25 \pm 1^\circ$ in the counter-clockwise direction compared to the lattice in the computational model. The lattice orientation was confirmed with two separate measurements — with a clean Cu surface and with an adsorbed (1S)-camphor molecule on the surface.

With this, we proceed to analyze the orientations and lengths of the image features and detect matching structures between AFM simulations and experiments.

D. Detecting matching structures

We combine our analysis of the observed features in the simulated and experimental AFM images and detect matches to identify the structures observed in AFM experiments. With each structure, we identify the orientations and lengths of the observed image features in each image in the stack (FIG. 2c). Due to the variation in the orientations and lengths at different heights, we compare the features via their average orientation and length over the stack of images. The orientations are measured as the angle from the nearest crystallographic axis of the Cu(111) surface in the clockwise direction, in the range $[0, 60]^\circ$. We measure the lengths as the distance between signal maxima. Images for all heights and the average lengths and orientations of all analyzed features and their standard deviations are provided in the Supplementary Material (SM).

In the simulated AFM images, the lengths of the main features vary from 2.76 to 4.04 Å (TAB. I). Their orientations show two distinct groups, in which structure A2 is nearly parallel (3.0°) to the crystallographic axis of the Cu(111) surface, and the other structures are near the middle region between the axes (on average 26.9°). In the experimental images, the main feature of structure E1 has noticeably shorter length (2.74 Å) than the other structures (on average 3.8 Å). Their orientations vary from 7.8 to 38.1°, with no distinct grouping in preferred directions.

The standard deviations of molecular orientations in the stacks of images are as large as 5° (TABS. SI and SII). The length of the image features in most cases increases by 0.1 Å for every Å that the tip is approaching the molecule. Consequently, the standard deviations of the feature lengths are large, up to 0.8 and 0.5 Å in the simulated and experimental images, respectively.

We match the structures using two identified features in structures A1, A3, E1, and E2, and one feature in structures A2 and E4. In this analysis, we detect 3 possible matches between structures E2-A3, E1-A1, and E4-A2 (FIG. 4). In E2 and A3, the orientations of the main and secondary features agree within 1.7 and 12.4°, respectively. The corresponding lengths agree within 0.5 and 0.6 Å. In E1 and A1, the orientations of the features agree within 13.0 and 5.1°, and the lengths within 0.7 and 0.1 Å, respectively. In E4 and A2, we compare a single feature, in which the orientation agrees within 10.8° and the length within 0.2 Å.

We also analyze how the image features in the simulated AFM images correspond to the atomistic model structures. For this, we measure the distance and orientation between the two topmost atoms of (1S)-camphor in the model structures A1–A4. Here, we observe two distinct groups of distances and orientations (TAB. I). In structures A1 and A4, the distance of 1.77 Å corresponds to the separation of H atoms in the same methyl group. Distances in structures A2 and A3 are considerably longer (2.23 and 2.54 Å, respectively) and originate from H atoms in different groups. The orientations exhibit two preferred directions: nearly parallel to the crystallographic axis (structure A2, 3.9°), and in the middle region between neighboring axes (structures A1, A3, and A4, on average 23.0°). The distance between the top atoms in the model structures is 1.4 Å shorter, on average, than the corresponding feature length in simulated AFM images. The average difference in the orientation between the model structures and simulated images is 3.1°, without a clear trend in either rotation direction.

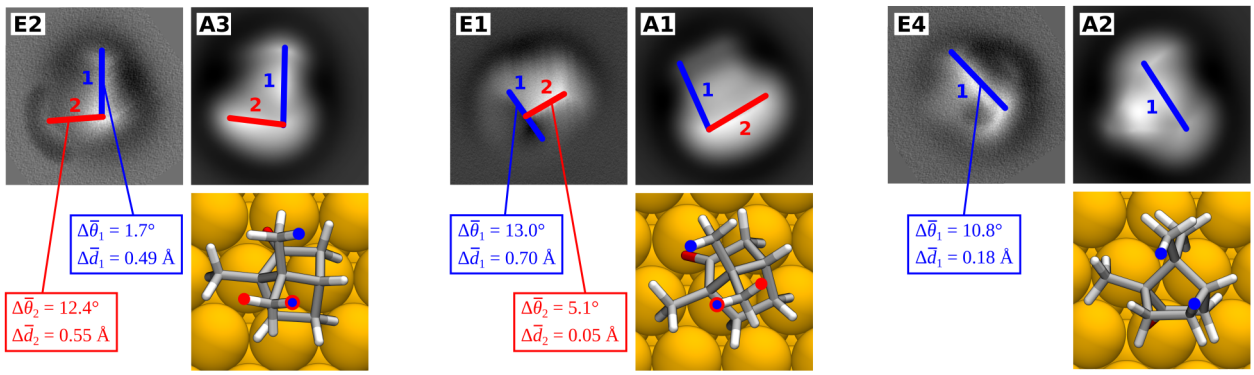

FIG. 4. **Detected matches between experimental and simulated structures.** Matches between the experimental (E) and simulated (A) structures E2-A3, E1-A1, and E4-A2 are compared via the orientations and lengths of the identified features 1 (blue) and 2 (red) in the AFM images. The top view of each simulated structure shows the topmost atoms (blue and red), which are the origin of the identified features. Matching accuracy is evaluated via the difference in the average orientations ($\Delta\bar{\theta}$) and lengths ($\Delta\bar{d}$) of the observed features between the experimental and simulated images.

III. DISCUSSION

The stable adsorbate structures, which we identified with BOSS and DFT, show that (1S)-camphor can adsorb to Cu(111) in multiple stable configurations with varying molecular orientations and adsorption sites. This explains the several types of adsorbates observed in AFM experiments. To interpret the experiments, we generated simulated AFM images of the most stable and least mobile model structures for a direct comparison with experiments using the image features.

In this comparison, we observed very similar features between experimental and simulated AFM images. We immediately detected 3 good matches between structures E2-A3, E1-A1, and E4-A2, in which the orientations and lengths of the observed features are in good agreement. In these matches, we also took into account the deviation of the features in each image stack.

The best match is between structures E2 and A3, in which the two analyzed image features agree closely between simulations and experiments. Similarly good agreement of two features was found between structures E1 and A1. Here, however, the image features emerge in different order as the CO tip is approaching the molecule. This can be explained by a minor tilt in the orientation of the molecule, which can be induced by its interaction with the tip. In the third match, between E4 and A2, the length of the single analyzed feature agrees closely and the orientation is only slightly outside the specified error threshold (within 11°).

In (1S)-camphor, the hydrogen atoms in the outer parts of the molecule are flexible and can vibrate or rotate once the CO tip is approaching the molecule [35]. However, no dynamic tip effects were included in BOSS or PP-AFM by method design. In future work, the accuracy of AFM simulations and the quality of structure matching could be further enhanced by implementing sample response to the presence of tip in the PP-AFM model.

In contrast to the 2D adsorbate case, 3D adsorbate AFM images contain features whose measured orientations and lengths exhibit much larger differences between theory and experiment, and thus require a thorough statistical analysis. We also note that the trend of increasing feature length with tip approach is opposite to what was reported for 2D molecules in full monolayer [37] or borders of polygons of C_{60} molecule [12]. These findings demonstrate the complexity of analyzing the structure of bulky and asymmetric molecular adsorbates, such as (1S)-camphor.

The adsorption of (1S)-camphor on Cu(111) has been previously studied by Alldritt et al. [35] using a neural network (NN) with image descriptors for automatic structure discovery. The configurations identified here match well with some of the higher probability configurations from the NN.

In conclusion, we have proposed a new approach to investigate the structure of complex 3D adsorbates. We have integrated a set of tools from different fields, using Bayesian inference enhanced structure search, AFM simulations with the PP-AFM model, and CO-AFM experiments. With BOSS, we constructed a surrogate model of the complete AEL to extract the stable model structures and their energy barriers of molecular mobility. This allowed us to infer different adsorbate types independently of AFM images, and free of chemical intuition. PP-AFM simulated images then facilitated a direct comparison of the model structures with CO-AFM experiments. The combination of findings derived from different sources is key to robust identification of distinct adsorbate geometries in experimental images. In the case of (1S)-camphor on the Cu(111) surface, we identified three different adsorbate geometries in the otherwise incomprehensible features of AFM experimental images. This Bayesian-based general approach can be applied to other adsorption structure search problems and combined with other experimental techniques. The subsequent knowl-

edge on adsorbate configurations can improve our control of the surface structure and help in tuning the properties of the materials.

IV. METHODS

A. Bayesian Optimization Structure Search

Stable structures of (1S)-camphor on Cu(111) were identified with BOSS in the minima of the 6D AEL [36]. The AEL was defined with respect to the position and orientation of the molecule using 3 translational and 3 rotational degrees of freedom. The Cu substrate was modeled with a 6×4 supercell with 4 atomic layers. The search range on the Cu(111) surface was within the orthogonal unit cell, with (1S)-camphor at height $z \in [4, 7]$ Å from the surface (measured from the top Cu layer to the center of the molecule). The search range of the 3 rotational degrees of freedom was $[0, 360]^\circ$.

The 6D AEL was predicted with 1380 energy points, which originated from 689 DFT calculations with applied symmetries in the orthogonal unit cell. The minima of the AEL were identified with BOSS using its local minima search functionality. The predicted structures were verified and the building block approximation validated with full relaxation in DFT (i.e. unrestricted motion of all atoms). The energy barriers of molecular diffusion (E_D) were evaluated from the 6D AEL model, and the rotational barriers (E_R) were predicted with BOSS by rotating the molecule in the relaxed structures.

Energies in sampling the AEL and structural relaxations were computed with DFT using the all-electron, numeric atom-centered orbital code FHI-aims [38, 39] with the Perdew-Burke-Ernzerhof (PBE) [40] exchange-correlation functional, augmented with van der Waals (vdW) corrections with the vdW$^\mathrm{surf}$ parametrization [41].

B. Simulated AFM

The PP-AFM simulations were based on DFT calculated geometries and electrostatic potentials [33]. For the mechanical part of PP-AFM we employed OPLS force-field [42] for the Lennard-Jones interactions and PP lateral stiffness of 0.24 N/m [37]. The PP was set 3 Å below the last metallic atom of the tip [43]. The electrostatic nature of the CO-tip was described via negative quadrupole on the PP with a moment of -0.025 e×Å^2 [44]. We used peak-to-peak amplitude of 1.0 Å for the conversion of forces to frequency shift Δf. All the heights refer to the center of the tip-oscillations.

We simulated a 3D stack of AFM images for heights in between 6.5 to 5.3 Å above the highest atom of the structures. The height is given with respect to the last metallic atom of the tip. The height step between each image was 0.1 Å. We also studied simulated images much closer than in the case of 2D molecules. This is due to hydrogen atoms that are responsible for the contrast in the AFM images. Hydrogen atoms evince Pauli repulsion much closer to the nucleus than carbons, which are important for the contrast of 2D molecules.

For each structure, we analyzed 11 images, in which we measured orientations (θ_sim) and lengths (d_sim) of the most pronounced lines. The Gwyddion program [45, 46] was used to find local maxima for the measurements. The lines of the measurements are marked in FIG. S1-S4. We also performed statistical analysis and linear fitting on the measured data. The results are presented in TAB. SI. The results for the most prominent lines (marked as 1 in FIG. S1-S4 and TAB. SI) are shown in TAB. I.

C. Experimental AFM

A polished Cu(111) single-crystal (Mateck/Germany) was prepared by repeated Ne+ sputtering (0.75 keV, 15 mA, 20 min) and annealing (850-900 K, 5 min) cycles. Surface cleanliness and structure was verified by scanning tunneling microscopy (STM). Sample temperatures during annealing were measured with a pyrometer (SensorTherm Metis MI16). (1S)-camphor (Sigma-Aldrich, purity > 98.5%) was introduced into the vacuum system via a leak valve and deposited onto the Cu(111) surface at a low-temperature ($T = 20$ K) to increase the number of distinct adsorption configurations and to achieve individual molecules rather than clusters on the surface.

The STM and CO-AFM images were taken with a Createc LT-STM/AFM with a commercial qPlus sensor with a Pt/Ir tip, operating at approximately $T = 5$ K in UHV at a pressure of 1×10^{-10} mbar. The quartz cantilever (qPlus sensor) had a resonance frequency of $f_0 = 29939$ Hz, a quality factor $Q = 101099$, and was operating with an oscillation amplitude $A = 50$ pm. Tip conditioning was performed by repeatedly bringing the tip into contact with the copper surface and applying bias pulses until the necessary STM resolution was achieved. The tip apex was functionalized with a CO molecule before AFM measurements. The STM images were recorded in constant-current mode, while the AFM operated in constant-height mode. Raw data was used as input for the image analysis. In order to minimize experimental artefacts that would cause problems with interpretation, we have implemented the following measures: Checking the background Δf before CO pickup (smaller value indicates sharper overall tip); scanning another CO to ensure the symmetry of the CO tip after tip passivization and prior to further AFM imaging; and confirming that the excitation (dissipation) signal remains flat/featureless during the AFM measurements.

D. Image analysis

We conducted an extensive image analysis to capture any statistical variation in the features of the experimen-

tal and simulated AFM images. Orientations and lengths of prominent features in experimental and computational datasets were determined by peak-to-peak analysis based on local maxima and minima in the AFM images. All orientation and length measurements were performed in Gwyddion [45, 46]. The results for the prominent features are available in TABS. SI and SII.

V. DATA AVAILABILITY

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

VI. CODE AVAILABILITY

The software used in this study is freely available, the BOSS code in [31], the Probe Particle Model in [34] and the Gwyddion in [46].

[1] P. Rotter, B. A. J. Lechner, A. Morherr, D. M. Chisnall, D. J. Ward, A. Jardine, J. Ellis, W. Allison, B. Eckhardt, and G. Witte, Coupling between diffusion and orientation of pentacene molecules on an organic surface, Nat. Mater. **15**, 397 (2016).

[2] K. Brandt, M. E. Chiu, D. J. Watson, M. S. Tikhov, and R. M. Lambert, Chemoselective catalytic hydrogenation of acrolein on Ag(111): Effect of molecular orientation on reaction selectivity, J. Am. Chem. Soc. **131**, 17286 (2009).

[3] W. Gao, Z. D. Hood, and M. Chi, Interfaces in heterogeneous catalysts: Advancing mechanistic understanding through atomic-scale measurements, Acc. Chem. Res. **50**, 787 (2017).

[4] E. D. Boyes, A. P. LaGrow, M. R. Ward, R. W. Mitchell, and P. L. Gai, Single atom dynamics in chemical reactions, Acc. Chem. Res. **53**, 390 (2020).

[5] L. Bartels, G. Meyer, and K.-H. Rieder, Controlled vertical manipulation of single CO molecules with the scanning tunneling microscope: A route to chemical contrast, Appl. Phys. Lett. **71**, 213 (1997).

[6] L. Gross, F. Mohn, N. Moll, P. Liljeroth, and G. Meyer, The chemical structure of a molecule resolved by atomic force microscopy, Science **325**, 1110 (2009).

[7] N. Pavliček, B. Fleury, M. Neu, J. Niedenführ, C. Herranz-Lancho, M. Ruben, and J. Repp, Atomic force microscopy reveals bistable configurations of dibenzo[a,h]thianthrene and their interconversion pathway, Phys. Rev. Lett. **108**, 086101 (2012).

[8] F. Albrecht, M. Neu, C. Quest, I. Swart, and J. Repp, Formation and characterization of a molecule–metal–molecule bridge in real space, J. Am. Chem. Soc. **135**, 9200 (2013).

[9] F. Albrecht, N. Pavliček, C. Herranz-Lancho, M. Ruben, and J. Repp, Characterization of a surface reaction by means of atomic force microscopy, J. Am. Chem. Soc. **137**, 7424 (2015).

[10] F. Albrecht, F. Bischoff, W. Auwärter, J. V. Barth, and J. Repp, Direct identification and determination of conformational response in adsorbed individual nonplanar molecular species using noncontact atomic force microscopy, Nano Lett. **16**, 7703 (2016).

[11] S. P. Jarvis, S. Taylor, J. D. Baran, N. R. Champness, J. A. Larsson, and P. Moriarty, Measuring the mechanical properties of molecular conformers, Nat. Commun. **6**, 8338 (2015).

[12] L. Gross, F. Mohn, N. Moll, B. Schuler, A. Criado, E. Guitián, D. Peña, A. Gourdon, and G. Meyer, Bond-order discrimination by atomic force microscopy, Science **337**, 1326 (2012).

[13] S. Kawai, T. Nishiuchi, T. Kodama, P. Spijker, R. Pawlak, T. Meier, J. Tracey, T. Kubo, E. Meyer, and A. S. Foster, Direct quantitative measurement of the C=O...H–C bond by atomic force microscopy, Sci. Adv. **3**, e1603258 (2017).

[14] F. Schulz, P. H. Jacobse, F. F. Canova, J. Van Der Lit, D. Z. Gao, A. Van Den Hoogenband, P. Han, R. J. Klein Gebbink, M.-E. E. Moret, P. M. Joensuu, I. Swart, and P. Liljeroth, Precursor geometry determines the growth mechanism in graphene nanoribbons, J. Phys. Chem. C **121**, 2896 (2017).

[15] S. Kawai, O. Krejčí, A. S. Foster, R. Pawlak, F. Xu, L. Peng, A. Orita, and E. Meyer, Diacetylene linked anthracene oligomers synthesized by one-shot homocoupling of trimethylsilyl on Cu(111), ACS Nano **12**, 8791 (2018).

[16] R. Pawlak, J. G. Vilhena, A. Hinaut, T. Meier, T. Glatzel, A. Baratoff, E. Gnecco, R. Pérez, and E. Meyer, Conformations and cryo-force spectroscopy of spray-deposited single-strand DNA on gold, Nat. Commun. **10**, 685 (2019).

[17] C. Moreno, O. Stetsovych, T. K. Shimizu, and O. Custance, Imaging three-dimensional surface objects with submolecular resolution by atomic force microscopy, Nano Lett. **15**, 2257 (2015).

[18] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, Phys. Rev. **136**, B864 (1964).

[19] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. **140**, A1133 (1965).

[20] M. Todorović, O. Stetsovych, C. Moreno, T. K. Shimizu, O. Custance, and R. Pérez, Pentacene/TiO_2 anatase hybrid interface study by scanning probe microscopy and first principles calculations, ACS Appl. Mater. Interfaces **10**, 34718 (2018).

[21] B. Schuler, G. Meyer, D. Peña, O. C. Mullins, and L. Gross, Unraveling the molecular structures of asphaltenes by atomic force microscopy, J. Am. Chem. Soc. **137**, 9870 (2015).

[22] S. Goedecker, Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems, J. Chem. Phys. **120**, 9911 (2004).
[23] F. Li, Y. Liu, L. Wang, J. Zhao, and Z. Chen, Improved stability of water clusters $(H_2O)_{30–48}$: A Monte Carlo search coupled with DFT computations, Theor. Chem. Acc. **131**, 1163 (2012).
[24] D. M. Packwood and T. Hitosugi, Rapid prediction of molecule arrangements on metal surfaces via Bayesian optimization, Appl. Phys. Express **10**, 065502 (2017).
[25] S. Carr, R. Garnett, and C. Lo, BASC: Applying Bayesian optimization to the search for global minima on potential energy surfaces, Proceedings of the 33rd International Conference on Machine Learning **48**, 10 (2016).
[26] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learning (MIT Press, Cambridge, Mass, 2006).
[27] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, Taking the human out of the loop: A review of Bayesian optimization, Proc. IEEE **104**, 148 (2016).
[28] M. Todorović, M. U. Gutmann, J. Corander, and P. Rinke, Bayesian inference of atomistic structure in functional materials, npj Comput. Mater. **5**, 35 (2019).
[29] A. T. Egger, L. Hörmann, A. Jeindl, M. Scherbela, V. Obersteiner, M. Todorović, P. Rinke, and O. T. Hofmann, Charge transfer into organic thin films: A deeper insight through machine-learning-assisted structure search, Adv. Sci. , 2000992 (2020).
[30] L. Fang, E. Makkonen, M. Todorović, P. Rinke, and X. Chen, Efficient cysteine conformer search with Bayesian optimization, arXiv:2006.15006 [physics.comp-ph] (2020).
[31] BOSS code repository, https://gitlab.com/cest-group/boss.
[32] P. Hapala, G. Kichin, C. Wagner, F. S. Tautz, R. Temirov, and P. Jelínek, Mechanism of high-resolution STM/AFM imaging with functionalized tips, Phys. Rev. B **90**, 085421 (2014).
[33] P. Hapala, R. Temirov, F. S. Tautz, and P. Jelínek, Origin of high-resolution IETS-STM images of organic molecules with functionalized tips, Phys. Rev. Lett. **113**, 226101 (2014).
[34] Probe Particle Model code repository, https://github.com/ProkopHapala/ProbeParticleModel.
[35] B. Alldritt, P. Hapala, N. Oinonen, F. Urtev, O. Krejčí, F. Federici Canova, J. Kannala, F. Schulz, P. Liljeroth, and A. S. Foster, Automated structure discovery in atomic force microscopy, Sci. Adv. **6**, eaay6913 (2020).
[36] J. Järvi, P. Rinke, and M. Todorović, Detecting stable adsorbates of (1S)-camphor on Cu(111) with Bayesian optimization, arXiv:2002.05598 [cond-mat.mtrl-sci] (2020).
[37] P. Hapala, M. Švec, O. Stetsovych, N. J. van der Heijden, M. Ondráček, J. van der Lit, P. Mutombo, I. Swart, and P. Jelínek, Mapping the electrostatic force field of single molecules from high-resolution scanning probe images, Nat. Commun. **7**, 11560 (2016).
[38] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, Ab initio molecular simulations with numeric atom-centered orbitals, Comput. Phys. Commun. **180**, 2175 (2009).
[39] V. Havu, V. Blum, P. Havu, and M. Scheffler, Efficient integration for all-electron electronic structure calculation using numeric basis functions, J. Comput. Phys. **228**, 8367 (2009).
[40] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. **77**, 3865 (1996).
[41] V. G. Ruiz, W. Liu, and A. Tkatchenko, Density-functional theory with screened van der Waals interactions applied to atomic and molecular adsorbates on close-packed and non-close-packed surfaces, Phys. Rev. B **93**, 035118 (2016).
[42] W. L. Jorgensen and J. Tirado-Rives, The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin, J. Am. Chem. Soc. **110**, 1657 (1988).
[43] M. Ellner, N. Pavliček, P. Pou, B. Schuler, N. Moll, G. Meyer, L. Gross, and R. Peréz, The electric field of CO tips and its relevance for atomic force microscopy, Nano Lett. **16**, 1974 (2016).
[44] B. de la Torre, M. Švec, G. Foti, O. Krejčí, P. Hapala, A. Garcia-Lekue, T. Frederiksen, R. Zbořil, A. Arnau, H. Vázquez, and P. Jelínek, Submolecular resolution by variation of the inelastic electron tunneling spectroscopy amplitude and its relation to the AFM/STM signal, Phys. Rev. Lett. **119**, 166001 (2017).
[45] D. Nečas and P. Klapetek, Gwyddion: an open-source software for SPM data analysis, Centr. Eur. J. Phys. **10**, 181 (2012).
[46] Gwyddion website, http://gwyddion.net/.

ACKNOWLEDGMENTS

We would like to thank A. S. Foster, F. Urtev, and N. Oinonen for fruitful discussion. O. K. would like to thank P. Hapala for help with the AFM simulations.

O. K. has been supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 845060. M. T. and P. R. have received funding from the Academy of Finland via the Artificial Intelligence for Microscopic Structure Search (AIMSS) project No. 13316601 the Flagship programme: Finnish Center for Artificial Intelligence FCAI. P. L. has received funding from the Academy of Finland via projects No. 311012 and 314882. J. J. has been funded by the Emil Aaltonen Foundation.

We gratefully acknowledge CSC – IT Center for Science, Finland, and the Aalto Science-IT project for generous computational resources.

AUTHOR CONTRIBUTIONS

J. J. identified the stable model structures. O. K. produced simulated AFM images of the model structures and B. A. imaged the experimental structures with AFM. O. K. and B. A. carried out the image analysis. M. T., P. L., and P. R. conceived the study and advised the work. All authors assisted in manuscript preparation.

COMPETING INTERESTS

The authors declare no competing interests.

MATERIALS & CORRESPONDENCE

Correspondence and requests for materials should be addressed to J. J.