



Detecting stable adsorbates of (1S)-camphor on Cu(111) with Bayesian optimization

Jari Järvi*, Patrick Rinke and Milica Todorović

Full Research Paper

Open Access

Address:

Department of Applied Physics, Aalto University, P.O. Box 11100, 00076 Aalto, Espoo, Finland

Email:

Jari Järvi* - jari.jarvi@aalto.fi

* Corresponding author

Keywords:

Bayesian optimization; camphor; Cu(111); density-functional theory; electronic structure; organic surface adsorbates; physical chemistry; structure search; surface science

Beilstein J. Nanotechnol. **2020**, *11*, 1577–1589.

<https://doi.org/10.3762/bjnano.11.140>

Received: 14 May 2020

Accepted: 16 September 2020

Published: 19 October 2020

This article is part of the thematic issue "Molecular assemblies on surfaces – towards physical and electronic decoupling of organic molecules".

Guest Editor: S. Maier

© 2020 Järvi et al.; licensee Beilstein-Institut.

License and terms: see end of document.

Abstract

Identifying the atomic structure of organic–inorganic interfaces is challenging with current research tools. Interpreting the structure of complex molecular adsorbates from microscopy images can be difficult, and using atomistic simulations to find the most stable structures is limited to partial exploration of the potential energy surface due to the high-dimensional phase space. In this study, we present the recently developed Bayesian Optimization Structure Search (BOSS) method as an efficient solution for identifying the structure of non-planar adsorbates. We apply BOSS with density-functional theory simulations to detect the stable adsorbate structures of (1S)-camphor on the Cu(111) surface. We identify the optimal structure among eight unique types of stable adsorbates, in which camphor chemisorbs via oxygen (global minimum) or physisorbs via hydrocarbons to the Cu(111) surface. This study demonstrates that new cross-disciplinary tools, such as BOSS, facilitate the description of complex surface structures and their properties, and ultimately allow us to tune the functionality of advanced materials.

Introduction

Current frontier technologies are increasingly based on advanced functional materials, which are often blends of organic and inorganic components. For example, in search for renewable energy solutions, hybrid perovskites are currently the best candidate to replace silicon in our solar cells [1]. In medicine, hybrid materials have been studied extensively for applications in tissue engineering [2] and drug delivery [3]. To optimize the

functional properties of these materials, we need detailed knowledge of their atomic structure. Of particular interest is the hybrid interface, which has a central role in defining the electronic properties of the material.

Assemblies of organic molecules on surfaces have been studied experimentally, for example with X-ray diffraction [4,5], scan-

ning tunneling microscopy [6-8] and atomic force microscopy (AFM) [9-11]. These methods have a considerable resolution in imaging planar surface structures, but interpreting images of bulky three-dimensional molecules on surfaces can be difficult, which prevents an accurate structure determination. In such cases, computations can help in detecting the most stable structures.

With atomistic simulations, we can determine the optimal structures by computing the potential energy surface (PES). We can identify stable structures in the minima of the PES and evaluate their mobility via the associated energy barriers. The most stable structure, that is the most probable structure in nature, corresponds to the global minimum of the PES. For its reliable identification, we must explore the PES thoroughly.

Calculating the full PES for complex hybrid materials requires either (i) fast energy computations, or (ii) an advanced method of constructing the complete PES with a small number of energy points. Classical force-field potentials are fast to compute, but they cannot accurately model hybrid materials, in which atomic interactions often feature a mixture of covalent and dispersive bonding, with charge transfer and polarization effects. Instead, we must employ quantum mechanical methods, such as density-functional theory (DFT) [12,13], for electronically accurate energy sampling. Regarding hybrid materials, this makes a thorough exploration of the PES prohibitively expensive with conventional phase-space exploration methods, such as minima hopping [14], Monte Carlo methods [15], or metadynamics [16], which typically require calculating thousands of energy points on the PES.

Traditionally, stable structures have been identified by initializing the minima search with estimated low-energy structures, based on chemical intuition [17,18], thus narrowing down the search space. With hybrid materials, however, this intuition is difficult to apply and can lead to biased or incorrect results. For example, with only partial knowledge of the PES, a metastable local minimum energy structure could easily be misinterpreted as the most stable global minimum.

Recently, Gaussian processes (GPs) [19] and Bayesian optimization (BO) [20] have been applied in modeling the PES to identify structures with minimum energy. GP regression has been used for example in local structure optimization [21], in finding minimum energy paths [22], and in predicting specific materials properties, such as melting temperature [23] or elasticity [24]. BO has been applied in detecting molecular conformers [25] and adsorbate structures [26,27], in identifying stable molecular compounds [28], and in discovering materials with low thermal hysteresis [29] or thermal conductivity [30].

Typically, previous studies have employed customized material-specific models, using, for example, a coarse-grained search space with discrete molecular configurations, or predetermined GP hyperparameters, at the cost of generality of the method.

In this work, we show that the recently developed Bayesian Optimization Structure Search (BOSS) machine learning method [31-34] provides a solution to the structure search conundrum. With BOSS, we adopt the aforementioned approach (ii) and construct the complete PES using a small number of energy points. To demonstrate the capabilities of BOSS, we apply it with DFT to the adsorption of (1S)-camphor ($C_{10}H_{16}O$, hereafter shortened as camphor) on the Cu(111) surface. Camphor is an exemplary case of a bulky molecule, which is difficult to image with microscopy. AFM experiments [35] have revealed various different conformers of camphor on Cu(111), which makes it ideal for benchmarking the BOSS method.

Our objective is to detect the stable adsorbate structures of camphor on Cu(111). With BOSS, we build a surrogate model of the PES of adsorption and data-mine this PES to identify the stable structures in its minima. We converge the model for a reliable detection of all the PES minima, not only the global energy minimum. We estimate the mobility of the adsorbates from the energy barriers extracted from the surrogate PES and analyze the electronic structure of each adsorbate. Our results provide insight into the adsorption of complex organic molecules on metallic substrates and pave the way to more complex studies of hybrid monolayer formation and hybrid interfaces.

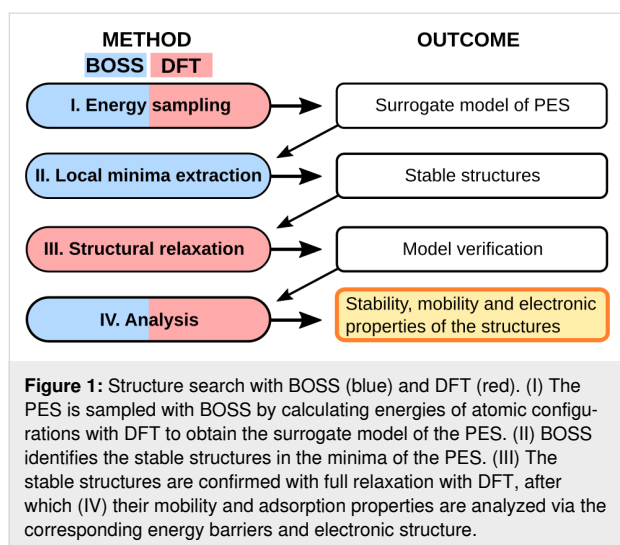
In the following sections, we first introduce our computational methods for adsorbate structure identification with BOSS, the first-principles calculations, and their application on detecting the stable adsorbates of camphor on Cu(111). We then present our results, discuss our findings, and conclude the analysis.

Computational Methods

Adsorbate structure identification

BOSS is a machine learning method that accelerates structure search via strategic sampling of the PES. With given initial data, BOSS builds the most probable surrogate model of the PES, refines it iteratively with active learning, and identifies the stable structures in the minima of the PES. In this work, we apply BOSS with DFT for accurate sampling of the energy points. In the following, we introduce the four-step process (Figure 1) of structure detection with BOSS and DFT, in analogy to [31]. We construct the surrogate model of the PES by sampling the adsorption energies with DFT (I). We then identify the stable structures by extracting the local minima of the PES (II) and verify them with full structural relaxation with DFT (III). We analyze the relaxed structures (IV) regarding

their stability and mobility via the energy barriers on the PES, and investigate their electronic properties with DFT.



Bayesian Optimization Structure Search

With the atomic structures and their corresponding energies, BOSS constructs a surrogate model of the PES. We define the atomic structures using chemical building blocks [36], which are natural rigid components of the structure, for example, rigid molecules, functional groups, or a surface slab. The PES is then defined in the phase space resulting from the remaining degrees of freedom, for example the relative translation and/or rotation of building blocks.

BOSS refines the PES model iteratively with active learning using BO (Figure 2a). We here only sketch the search principle and refer the interested reader to a more in-depth presentation and to the theoretical foundation in [19,31,37]. BO is a two-step process, in which data is first fitted with a GP distribution over functions using Bayesian regression. With the resulting

surrogate model (Figure 2b), BOSS then determines the next sampling point using an acquisition function.

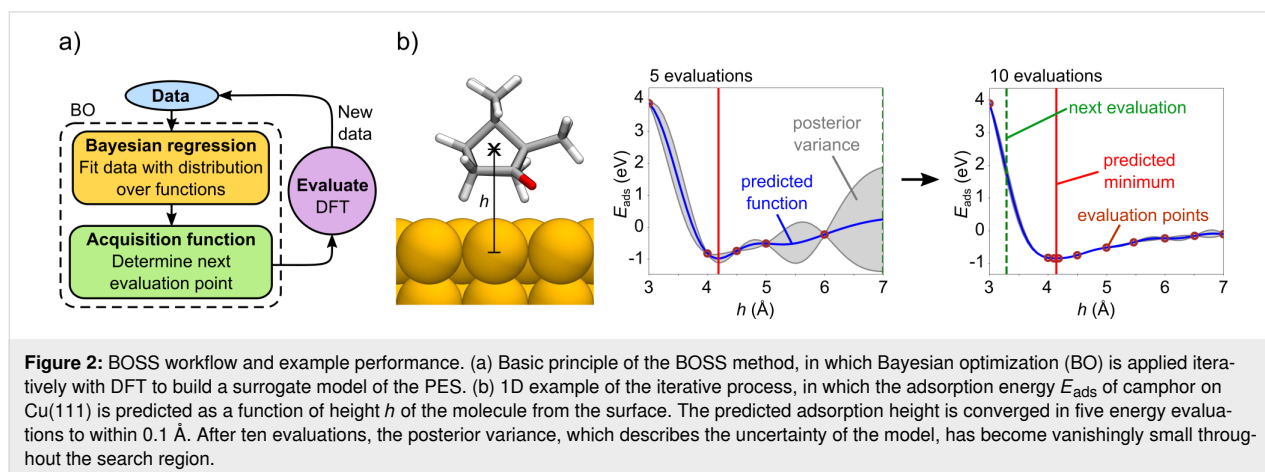
In the surrogate model, the posterior mean is the most probable model of the predicted function (here the PES). The posterior variance describes the uncertainty of the model in less explored areas. It therefore vanishes at the known data points.

The next sampling point is determined using the exploratory Lower Confidence Bound (eLCB) [38] acquisition function, which balances exploitation against exploration. In exploitation, BO refines the model by acquiring the next point near the currently predicted global minimum. In exploration, the next acquisition is made at the point of maximum posterior variance, exploring less visited areas. In this study, we converge the PES model with respect to the coordinates and energy of all the minima, not only the global energy minimum.

Local minima and barrier extraction

Once the PES is converged, we data-mine the surrogate model. We extract the lowest energy minima, which we equate with the lowest-energy adsorbate structures. The minima are detected using the built-in local minima search functionality of BOSS. The search is performed with minimizers, which apply the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) [39] optimization algorithm. The minimizers start in different regions of the PES and traverse the landscape, following the gradients to locate the minima.

The confidence of the surrogate model in different regions of the PES is quantified via the standard deviation (σ^B), which is the square root of the posterior variance in the GP model (Figure 2b). With the standard deviation, we evaluate the confidence of the surrogate model in the identified minima. Furthermore, we evaluate the accuracy of the model by computing the energy of each identified minima structure with DFT (E^D) and



compare it to the corresponding energy in the surrogate model (E^B).

The BOSS PES also provides access to energy barriers, with which we can estimate the mobilities of our identified adsorbate structures. BOSS provides post-processing tools to locate the lowest energy barriers between two minima in the PES with the nudged elastic band (NEB) method. Since we compute the PES with the building block approximation and not in the space of all atomic degrees of freedom, these energy barriers are only upper limits to the true barriers. However, even qualitative accuracy in barrier evaluations suffices to identify the least mobile structures, which are the best candidates when compared to experimentally observed structures. We will return to energy barriers and our way of estimating them in the Results section, after we have introduced the camphor/Cu(111) system in more detail.

Structural relaxation and analysis

We verify the identified structures against a full DFT structure relaxation. In this, we remove the building block approximation and allow unrestricted motion of all atoms according to the interatomic forces in DFT. We then quantify and analyze the structural changes in the relaxation with respect to the atomic coordinates and the energy change (ΔE_R^D) for each structure.

To validate the building block approximation, we evaluate the changes in the internal geometry of the building blocks after releasing them in the relaxation. For this, we calculate the average root-mean-square deviation of the atomic positions and the mean deviation of bond lengths, comparing the structures before and after the relaxation.

We furthermore investigate the electronic structure of the stable adsorbates by analyzing their partial density of states (DOS)

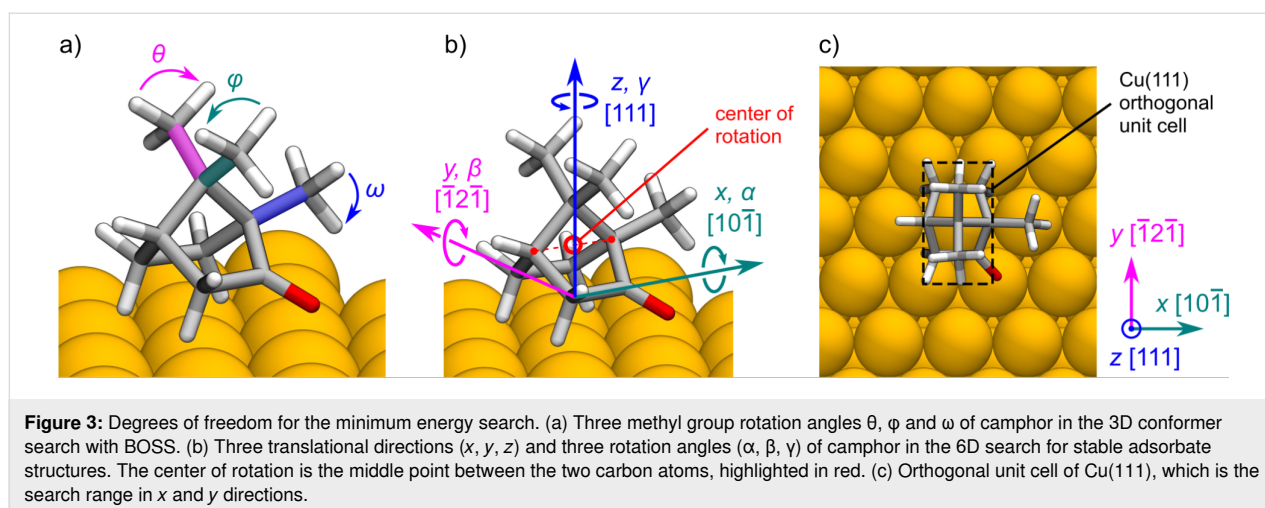
and the charge distribution with the Mulliken analysis of partial charges [40].

Camphor on Cu(111)

We study the adsorption of camphor on the Cu(111) surface using two building blocks: (i) the global minimum camphor conformer and (ii) the Cu(111) surface slab. With BOSS, we first identify the global minimum camphor conformer without the Cu(111) surface with a 3D search of methyl group rotations (Figure 3a). We normalize the lengths and angles of the C–H bonds in the three methyl groups by setting them to identical values, based on their average lengths and angles (see Supporting Information File 1). With this, we obtain an ideal camphor geometry with three identical minima in the methyl group rotation (i.e., 120° periodicity). We then study the rotation of the methyl groups in the ranges $\theta, \varphi, \omega \in [-60, 60]^\circ$.

With the identified global minimum conformer, we study the adsorption of camphor on Cu(111) with respect to molecular orientation and location. We define the PES of adsorption in a 6D phase space with three rotational angles (α, β, γ) and three translational directions (x, y, z), which correspond to the Cu lattice directions $[10\bar{1}]$, $[\bar{1}2\bar{1}]$ and $[111]$, respectively (Figure 3b). The adsorption height of the molecule (z) is defined with respect to the center point of rotation (Figure 3b), which is the middle point of the line connecting two C atoms at the sides of the rigid cage of camphor. We investigate the orientation of the molecule with full 360° rotation of all angles, in the range $(\alpha, \beta, \gamma) \in [-180, 180]^\circ$. The search range in the x - y plane of the Cu(111) surface is $(x, y) \in [-0.5, 0.5]$ (Figure 3c), defined in fractional unit cell coordinates, which corresponds to lattice vectors $[a', b'] = [2.57, 4.45] \text{ \AA}$.

Before we embarked on the full 6D camphor-on-Cu(111) search, we first scanned the system with several low-dimen-

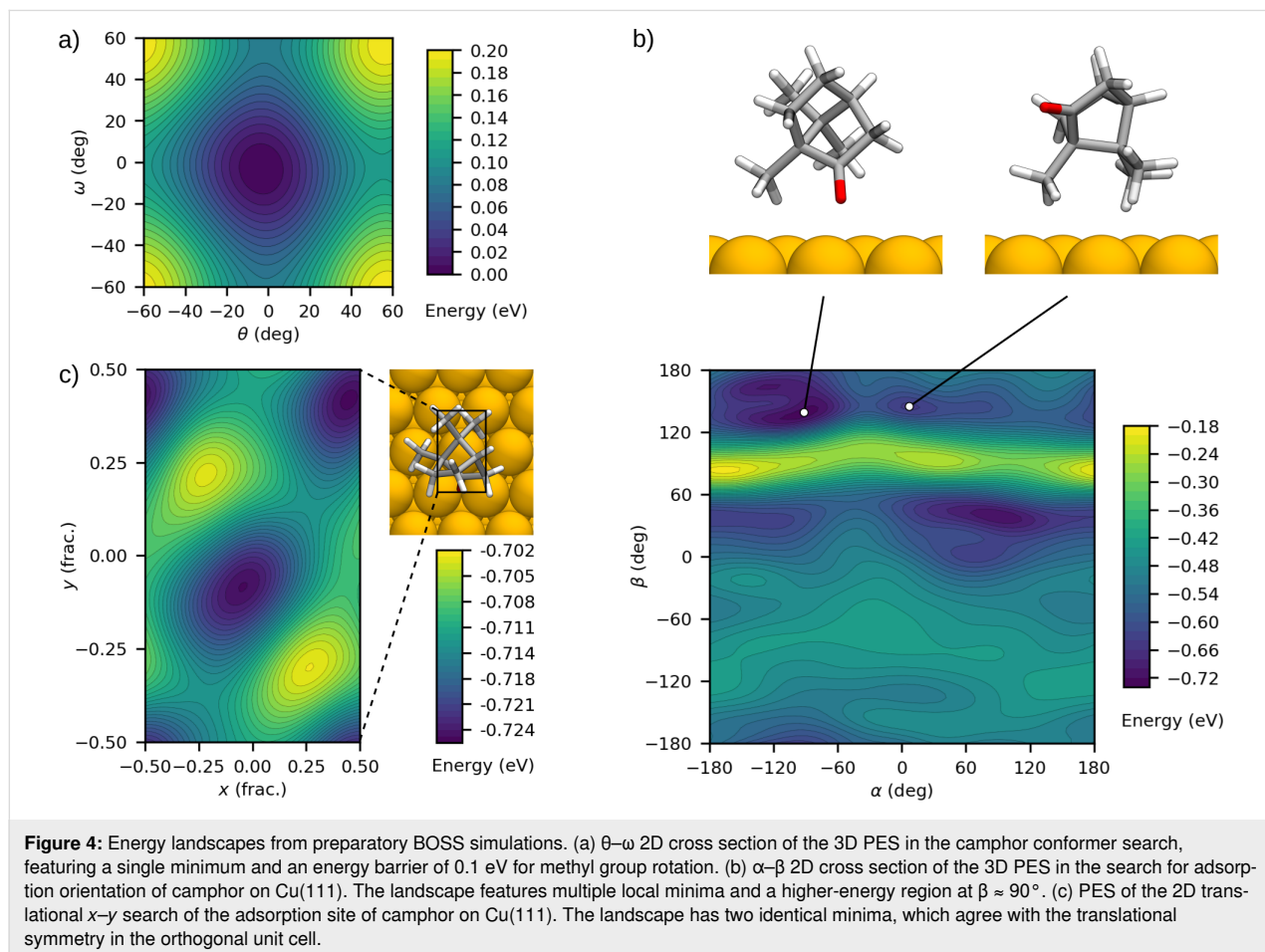


sional searches. Such low-dimensional searches (e.g., 1D variation of the adsorption height or 2D scans of molecular registry on the surface) permit us to relatively quickly explore the behavior of the system. We use them to find appropriate limits for the search dimensions (e.g., maximum and minimum height over the surface). Additionally, low-dimensional simulations help us to assess the contributions from rotational and translational degrees of freedom separately, to estimate the expected number of local minima and their approximate values, and to develop qualitative checks for expected energy landscapes (e.g., reflecting surface symmetries). The computational effort associated with these preparatory simulations is recycled, since all points sampled in reduced dimensions later serve as input in the 6D study. We note that analysis of low-dimensional simulations only provides us with qualitative insight into surface adsorption. Quantitative conclusions on the stable structures can only be drawn from a full 6D search.

With BOSS, we perform three low-dimensional searches, in which we study the adsorption of camphor on Cu(111) as a function of its (i) adsorption height (1D), (ii) orientation (3D), and (iii) adsorption site (2D). First, we investigate the height of

the molecule with a 1D search (Figure 2b) to determine a suitable height for the rotational search. Based on the resulting energy curve we estimate the optimal height at which we avoid high-energy peaks in all molecular orientations, and conduct the 3D rotational search. We then set the molecule in the observed minimum energy orientation (Figure 4b) and perform a 2D search of the adsorption site within the orthogonal unit cell of the Cu(111) surface (Figure 3c). With the acquired knowledge of the energy ranges, we then determine the optimal height range of the molecule for the 6D search.

We perform a 6D search with combined degrees of freedom to identify the stable adsorbate structures of camphor on Cu(111). The search is initialized using the previously acquired energy points from the low-dimensional studies. We multiply the number of initial energy points by applying the twofold translational symmetry in the orthogonal unit cell and the threefold rotational symmetry of the Cu(111) surface at the top site. With BOSS, we acquire new energy points and converge the 6D PES with respect to the energy and coordinates of the identified local minima (details provided in Supporting Information File 1).



The electronic structure of the stable adsorbates is analyzed with the partial DOS and the Mulliken analysis of partial charges. We compare the partial DOS of the adsorbed camphor to the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO, respectively) of an isolated camphor molecule. In the Mulliken analysis, we calculate the sum of partial atomic charges per element in the adsorbed camphor and compare them to the corresponding charge distribution of an isolated molecule. With this analysis, we study the effect of adsorption on the electronic structure of camphor in the identified stable structures.

First-principles calculations

We use density-functional theory to calculate the adsorption energy of camphor on Cu(111) in the BOSS runs, to relax the predicted stable structures and to analyze the electronic structure of the stable adsorbates. We apply the all-electron, numeric atom-centered orbital code FHI-aims [41-43] with the Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional [44]. PBE is augmented with van der Waals (vdW) corrections employing the vdW^{surf} parametrization [45] of the Tkatchenko–Scheffler method [46]. Previous work found that PBE + vdW^{surf} adequately describes organic molecules on metal surfaces [45,47,48].

Our converged settings employ tier-1 basis sets with light grid settings and a Γ -centered $3 \times 2 \times 1$ k -point mesh with a $(6 \times 4)\sqrt{3}$ supercell model. We apply relativistic corrections with the zero-order regular approximation [49] and Gaussian broadening of 0.1 eV of the electronic states. The total energy is converged within 10^{-6} eV in the self-consistency cycle and the structures are relaxed below a maximum force component of 10^{-2} eV/Å.

We model the Cu substrate with a Cu(111) slab of four atomic layers and $(6 \times 4)\sqrt{3}$ orthogonal unit cells (192 atoms, lattice vectors $[a, b, c] = [15.41, 17.79, 56.29]$ Å). The lattice constant of Cu is set to 3.632 Å, which we obtain from relaxed bulk Cu, in agreement with reference studies [50,51]. We construct the four-layer Cu slab by fixing the two bottom layers to their optimal layer separation ($d_{34} = 2.097$ Å, corresponding to bulk Cu). The two top layers are then relaxed, which results in a reduced layer separation ($d_{12} = 2.076$ Å, $d_{23} = 2.081$ Å), in agreement with previous calculations [52]. We apply this Cu slab model as a building block in the subsequent study of camphor adsorption.

Our other building block is the global minimum conformer of camphor, which we add onto the Cu slab model. The $(6 \times 4)\sqrt{3}$ supercell provides a good approximation of a single molecule on the surface, with an average lateral separation of 10 Å be-

tween the periodic images of camphor and 50 Å separation between the periodic Cu(111) slabs.

The adsorption energy E_{ads} is calculated as

$$E_{\text{ads}} = E_{\text{tot}} - (E_{\text{Cu}} + E_{\text{cam}}), \quad (1)$$

in which E_{tot} is the total energy of the camphor/Cu(111) system, E_{Cu} is the energy of the relaxed Cu slab, and E_{cam} is the energy of an isolated camphor molecule.

Results

Camphor conformer search

We analyzed the camphor conformers with a 3D BOSS search of the methyl group rotations. The energy landscape (Figure 4a), converged in 20 evaluations, features a single global energy minimum at $(\theta, \varphi, \omega) = (-3, 7, -3)^\circ$, and an energy barrier of 0.1 eV for the rotation of the methyl groups. Given this barrier, the rotation of the methyl groups $\Delta\varphi$ away from the global minimum is expected to be small at room temperature. The Arrhenius law predicts that in 50% of the molecules $\Delta\varphi < 10^\circ$, and in 70% $\Delta\varphi < 15^\circ$. Camphor is likely to be found in a conformation very close to the global minimum geometry. We thus take the identified conformer as a building block in the following adsorption study. Any further structural deformations are accounted for at a later stage with full DFT relaxation.

Qualitative insight into adsorption of camphor on Cu(111)

Before conducting a full 6D search, we carried out several low-dimensional searches to develop a feeling for the behavior of camphor on Cu(111). First, we performed a 1D search in the z direction, then a 3D rotational search in (α, β, γ) , and finally a 2D translational search in the x - y plane.

We learned about the adsorption height range of camphor on Cu(111) from a 1D BOSS search (Figure 2b) within the limits $z \in [3, 7]$ Å (other variables were set to $(x, y, \alpha, \beta, \gamma) = (0, 0, 0, 0, 0)$). The predicted minimum of the adsorption energy converged in five evaluations and is found at -0.847 eV at a height $z = 4.14$ Å. The energy curve has a strong dispersive character and the repulsive energy increases rapidly as the molecule approaches the surface below 4 Å.

For the 3D rotational study, we placed the molecule into a fixed position at $(x, y, z) = (0, 0, 5)$ Å to avoid close contact between the molecule and the surface. Molecular placement at the top site (above a Cu atom) here allows us to curtail the γ range to $[-60, 60]^\circ$. The resulting PES (Figure 4b) converged in 115

evaluations and contains many features associated with different reactive sites of camphor. The higher energy band at $\beta \approx 90^\circ$ corresponds to the closest approach of the molecule to the surface (via methyl group ω in Figure 3a). The multiple minima and strong barriers imply that camphor may adsorb on Cu(111) in various stable orientations. We explored the structures associated with the most favourable minima to infer the binding mechanisms. As shown in Figure 4b, we found that both charge-withdrawing O and neutral methyl groups face the surface, suggesting that both chemical and dispersive bonding can be expected in the full 6D search.

The 2D search in the x - y plane allowed us to compute the translational energy landscape for camphor. We set the molecule to the global minimum orientation $(\alpha, \beta, \gamma) = (-84, 143, 3)^\circ$ from the previous rotational search, at $z = 5 \text{ \AA}$. The PES (Figure 4c) converged in 20 evaluations and features two identical minima at $(x, y) = (-0.05, -0.08)$ and $(0.45, 0.42)$ in fractional coordinates of the unit cell. These correspond to the translational symmetry of the Cu(111) surface in the orthogonal unit cell. We conclude that our model fitting is qualitatively correct even when the landscapes are very flat, as with this choice of parameters. The flat energy landscapes indicate that rotational degrees of freedom may influence adsorption more than translational ones, but this is best verified in 6D.

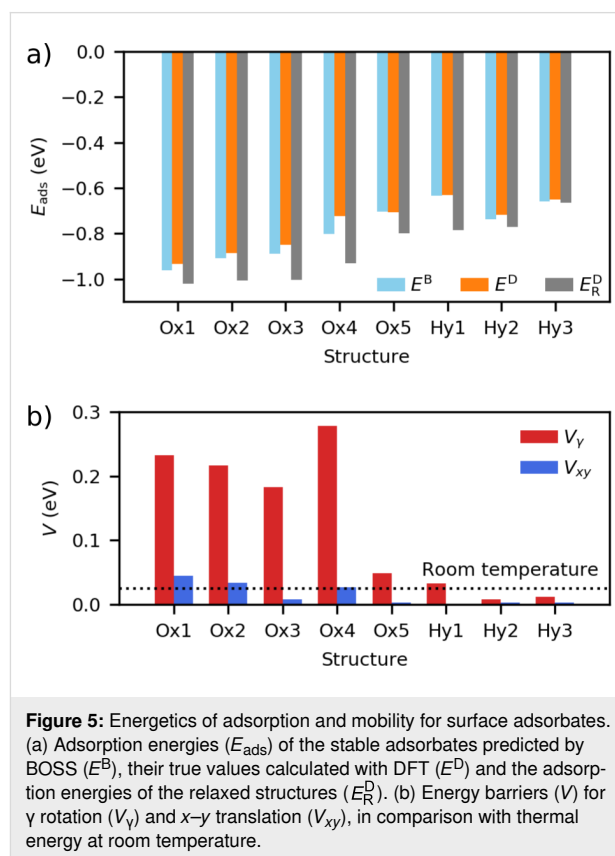
Based on the low-dimensional studies, we expect to find multiple stable adsorbate structures in the 6D search, with varying molecular orientations and both chemical and dispersive bonding. Given the observed energy ranges, we conclude that the optimal search range for the height of the molecule in the 6D search is $z \in [4, 7] \text{ \AA}$. The range is sufficiently broad to include all the minima and avoids high-energy peaks in the closest approach of the molecule to the surface.

Predicted stable adsorbates

For the 6D search of stable adsorbates, we employed the 492 previously acquired energy points from the low-dimensional studies. These points were then multiplied according to the translational and rotational symmetries of the Cu(111) surface, which resulted in 986 initial energy points for the 6D search. We converged the 6D PES (details provided in Supporting Information File 1) by acquiring 197 new points, for which we also applied the symmetries. The surrogate model of the 6D PES was then constructed with 1380 energy points.

In the minima of the PES, we identified eight unique stable structures with predicted adsorption energies (E^B) in the range $[-0.961, -0.634] \text{ eV}$ (Figure 5a and Table 1). We have classified the structures with respect to the bonding species closest to the surface in the adsorbed camphor, namely oxygen (class Ox)

and hydrogen (class Hy). The standard deviation of the adsorption energy (σ^B) in the surrogate PES is 0.019 eV in the global minimum and 0.025 eV on average over all minima (Table 1), which shows low uncertainty of the model in these points. The energies of the identified structures, calculated with DFT (E^D) are in the range $[-0.933, -0.631] \text{ eV}$, in close agreement with the predicted energies.



Relaxed structures

We verified the stable structures by performing full DFT relaxations (Figure 6a,b). In the relaxation, we observed an average decrease of -0.11 eV from the E^D energies (Figure 5a and Table 1). We found that in class Ox structures, 80% of the binding energy is due to dispersion whereas in class Hy structures the binding energy is purely dispersive.

The structural changes in the relaxation were analyzed by comparing the location and orientation of the molecule before and after the relaxation. We observed the relaxed structures to be almost identical with the initial ones. The average change in the location of the molecule, over all structures, is $(\overline{\Delta x}, \overline{\Delta y}, \overline{\Delta z}) = (0.13, 0.09, 0.19) \text{ \AA}$ and in the orientation $(\overline{\Delta \alpha}, \overline{\Delta \beta}, \overline{\Delta \gamma}) = (6.1, 5.8, 2.5)^\circ$. The structural changes in the Cu slab are minimal. The changes in the internal geometry of camphor in the relaxation, after removing the building block

Table 1: Adsorption energies of the stable adsorbates, predicted by BOSS (E^B), and their standard deviation in the surrogate model of the 6D PES (σ^B). Adsorption energies calculated with DFT (E^D) and their difference from the predicted energies (ΔE^D). Energy after relaxation (E_R^D), and energy change in the relaxation (ΔE_R^D). Predicted energy barriers of γ rotation (V_γ) and x - y translation (V_{xy}).

	E^B (eV)	σ^B (eV)	E^D (eV)	ΔE^D (eV)	E_R^D (eV)	ΔE_R^D (eV)	V_γ (eV)	V_{xy} (eV)
Ox1	-0.961	0.019	-0.933	+0.028	-1.022	-0.089	0.232	0.045
Ox2	-0.910	0.013	-0.885	+0.025	-1.008	-0.123	0.216	0.034
Ox3	-0.889	0.027	-0.850	+0.039	-1.005	-0.155	0.183	0.008
Ox4	-0.803	0.032	-0.723	+0.079	-0.932	-0.209	0.278	0.027
Ox5	-0.704	0.016	-0.706	-0.002	-0.800	-0.094	0.048	0.003
Hy1	-0.634	0.021	-0.631	+0.003	-0.784	-0.154	0.033	0.001
Hy2	-0.737	0.041	-0.719	+0.019	-0.772	-0.053	0.008	0.003
Hy3	-0.658	0.027	-0.652	+0.005	-0.664	-0.012	0.012	0.003

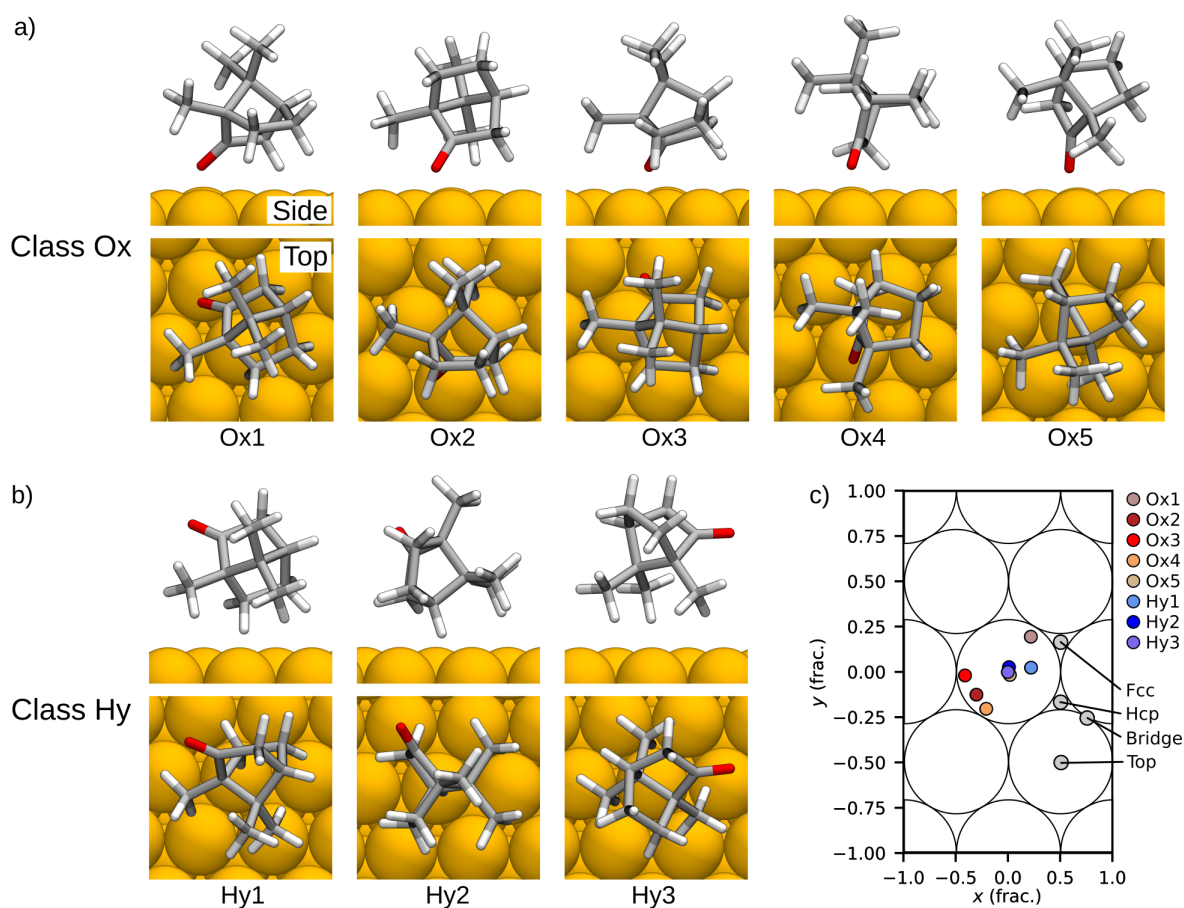


Figure 6: Relaxed stable adsorbate structures of camphor on Cu(111) in the 6D search, showing (a) chemisorption of the molecule via oxygen (class Ox) and (b) physisorption via hydrogen (class Hy). (c) Adsorption site of camphor in the relaxed structures (center of the molecule) and the high-symmetry points of the Cu(111) surface.

approximation, were evaluated using the average root-mean-square deviation of the atomic positions and the mean deviation of bond lengths, which are 0.13 Å and 0.003 Å, respectively, on average over all structures (see Supporting Information File 1 for structure-specific data).

We analyzed the adsorption site of camphor in the relaxed structures (Figure 6c) with respect to the center of the molecule (Figure 3b). The adsorption sites show a notable difference between the two classes. Class Hy structures (in particular Hy2 and Hy3) prefer to adsorb close to the top site, whereas class Ox

structures feature more variance in their location. Three of the class Ox structures (Ox1, Ox3, and Ox4) adsorb near the bridge site and Ox5 is close to the top site.

To estimate the mobility of camphor molecules on the surface, we inspect translational and rotational barriers. The translational energy barriers were computed using 2D x - y cross sections (grid of 100×100 points) of the predicted 6D PES, as described in the Computational Methods section. For the γ rotation barriers, we extracted 1D γ energy profiles from the 6D PES but found them overly smooth and free of features expected for an asymmetric molecule rotating on the Cu(111) surface. We concluded the upper limits for γ rotation to be too inaccurate and analyze the γ energy barriers using the fully relaxed structures of local minima geometries. For each minimum type, we rotated camphor in-place (center point of rotation in Figure 3b) and computed the rotational energy profile with a 1D BOSS search (converged in 15 evaluations). While this approach is still approximate, the resulting energy profiles exhibit features that correctly reflect surface symmetry and provide us with a better estimate of the barriers without investing time and computational expense into NEB calculations.

The predicted energy barriers of γ rotation and x - y translation (Figure 5b and Table 1) are in the range [0.008, 0.278] and [0.001, 0.045] eV, respectively. The barriers are highest in class Ox structures, specifically in structures Ox1–Ox4, with a notable difference to class Hy. When we take into account the standard deviation of the adsorption energy in the surrogate model (Table 1), the smallest energy barriers (of the order of 0.01 eV and below) are practically zero. This indicates free rotation of structures Hy2 and Hy3, and free diffusion of structures Ox3, Ox5, and Hy1–Hy3, even at low temperatures.

Electronic structure

We analyzed the charge distribution of the stable adsorbates with the Mulliken analysis of partial charges and investigated their partial DOS to study the effect of adsorption on the electronic structure. The Mulliken analysis of partial charges in the relaxed structures (Figure 7a and Table 2), in comparison to the charge distribution of an isolated camphor molecule, shows electron transfer from the adsorbed camphor molecule to the Cu substrate. The electron transfer is highest in class Ox structures, in which the O atom of camphor is close to the Cu surface. The average partial charge of camphor (Δq) is $+0.21 e$ (elementary charge, $e = |e^-|$) in class Ox structures and $+0.10 e$ in class Hy structures. In class Ox structures, the main contribution to the positive charge comes from hydrogen (H) atoms, with O as the second notable contributor. In class Hy structures, the positive charge of camphor originates predominantly from H atoms.

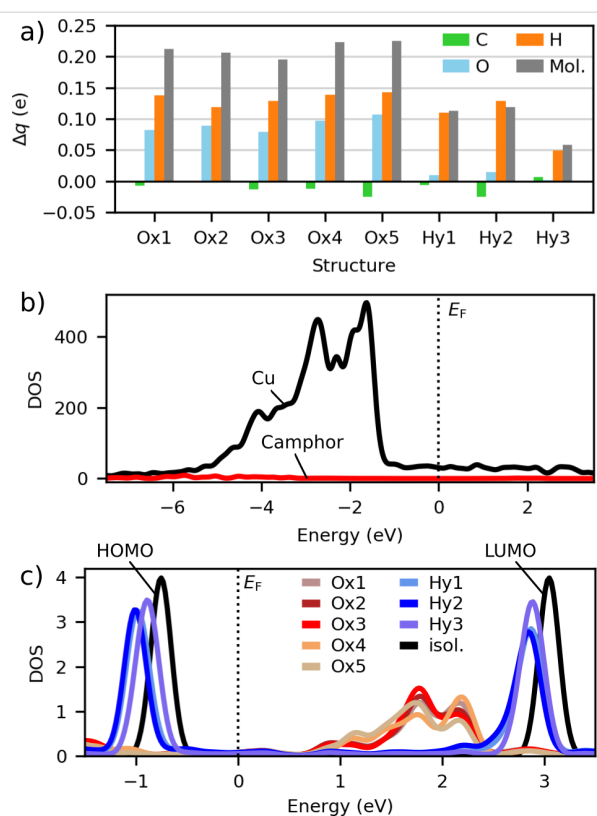


Figure 7: Electronic properties of different camphor adsorbates. (a) The sum of partial charges (Δq) in the adsorbed camphor in the relaxed structures. (b) DOS of Cu and camphor in structure Ox1, and (c) DOS of camphor in the relaxed structures and in an isolated molecule.

Table 2: The sum of partial charges of C, O, and H in an adsorbed camphor molecule (Δq_C , Δq_O , and Δq_H , respectively) and the total partial charge of the camphor molecule (Δq).

	Δq_C (e)	Δq_O (e)	Δq_H (e)	Δq (e)
Ox1	-0.01	+0.08	+0.14	+0.21
Ox2	-0.00	+0.09	+0.12	+0.21
Ox3	-0.01	+0.08	+0.13	+0.20
Ox4	-0.01	+0.10	+0.14	+0.22
Ox5	-0.02	+0.11	+0.14	+0.23
Hy1	-0.01	+0.01	+0.11	+0.11
Hy2	-0.03	+0.01	+0.13	+0.12
Hy3	+0.01	+0.00	+0.05	+0.06

In the partial DOS of the relaxed structures (Figure 7b,c), we analyze the electronic states of the adsorbed camphor close to the Fermi level. The partial DOS of class Ox structures features hybridization of the electronic states, in comparison to the HOMO and LUMO of an isolated camphor. The hybridization implies chemical bonding between the molecule and the substrate in class Ox. Conversely, in class Hy, the electronic states

resemble the HOMO and LUMO of an isolated camphor molecule and are at -1.0 and 2.9 eV, respectively, with an energy gap of 3.9 eV. This indicates physisorption between the molecule and the substrate in class Hy.

Discussion

With the low-dimensional studies of molecular translation (1D and 2D) and rotation (3D), we obtained a qualitative description of the adsorption properties of camphor on Cu(111). We gained insight into the estimated adsorption height of the molecule and acquired the ranges of adsorption energy with respect to molecular orientation and the adsorption site. The rotational energy landscape with multiple local minima suggests that camphor can adsorb on Cu(111) in various stable orientations. From the low-dimensional analysis, we obtained the required knowledge to determine the optimal search range for the height of the molecule in the subsequent 6D study.

In the relaxation of the identified structures, we observed minor changes in the molecular orientation, the adsorption site, and the adsorption energy. This effectively confirms the accuracy of the surrogate model of the 6D PES. Negligible changes in the internal structure of camphor and the Cu slab in the relaxation validates the building block approximation in this study.

The eight stable adsorbate structures extracted from the 6D search feature notable differences between the class Ox and class Hy structures, specifically, regarding their adsorption energy, adsorption site, energy barriers, and the electronic structure. Class Ox adsorbates have the highest adsorption energies and high energy barriers of molecular mobility. In class Ox structures, the preferred adsorption site is near the bridge site, so that the O atom can point sideways to bond with the Cu atom. In class Hy structures, methyl groups avoid the top site, so the molecule centers there, and the methyl groups point sideways. The DOS of class Ox structures feature hybridization of the electronic states and the electron transfer from the molecule to the substrate is significantly larger than in class Hy structures, with the largest contribution per atom from O. This indicates chemisorption of camphor via O to the Cu substrate. Conversely, in class Hy structures we observed the characteristics of physisorption. Class Hy structures have systematically lower adsorption energies, energy barriers, and electron transfer to the substrate, and their DOS resembles the HOMO and LUMO of an isolated molecule. These findings are supported by the vdW contributions in the adsorption energy, which show 80% dispersive bonding in class Ox structures and fully dispersive adsorption in class Hy structures.

To verify the identified stable structures, we can compare them to adsorbates observed in experiments. The adsorption of

camphor on Cu(111) has been studied experimentally with AFM by Alldritt and co-workers [35]. In their images, they have observed various different adsorbate structures, which shows that camphor can adsorb on Cu(111) in multiple stable configurations. In the experiments, camphor molecules were deposited onto the Cu surface at 20 K temperature and the imaging was done at 5 K. When the surface is annealed to the imaging temperature, we expect the deposited molecules to obtain the global minimum conformer geometry, which corresponds to the camphor building block in this study. Based on the estimated energy barriers of molecular mobility in this study, we conclude that the experiments likely feature chemisorbed camphor molecules from class Ox. In particular, the structures Ox1–Ox4, which have the highest barriers, are the most likely candidates for static adsorbates. They also have the highest adsorption energies, which makes them the most probable structures to be observed. Conversely, class Hy structures, which have lower adsorption energies and low energy barriers for molecular mobility, are less likely to be imaged in experiments. A more detailed comparison between BOSS and AFM will be reported in [53].

We highlight the computational efficiency of global structure search with BOSS by comparing the number of required DFT calculations to a conventional structure search. The best candidates for the minimum-energy structures can be first estimated using chemical intuition and then relaxed with DFT to identify the stable structures. With camphor on Cu(111), we can search for the stable adsorbates by placing the molecule on each of the four high-symmetry points of the Cu surface (Figure 6c) and investigate, for example, ten different molecular orientations at each of the adsorption sites. We estimate that the relaxation of the structures requires on average 40 calculation steps per structure. With this method, the estimated computational cost would be 1600 DFT calculations. Still, this amounts to exploring only a small portion of the PES and does not guarantee a reliable identification of the global minimum energy structure.

With BOSS, we identified the stable structures of camphor on Cu(111) with 892 DFT calculations (689 to construct the surrogate model of the 6D PES, and 203 to relax the eight structures). Relaxation of the predicted stable structures in the local minima of the PES was fast (25 relaxation steps per structure on average) due to their low initial energy. With the PES model, we were able to reliably identify not only all the minima, but also the associated energy barriers of molecular mobility. This comparison highlights the benefits of the BOSS approach, which are, in particular, (i) computational efficiency, (ii) reliable identification of the most stable structures, and (iii) energy barriers readily obtained with the surrogate model of the PES.

Conclusion

In this study, we have demonstrated the efficiency of BOSS in global structure search with complex molecular adsorbates. We have shown the accuracy of the constructed surrogate model of the PES, in comparison with adsorption energies of stable structures calculated with DFT. As a benchmark system, we have analyzed the adsorption of a camphor molecule on the Cu(111) surface with respect to molecular translation and rotation. With BOSS, we constructed a surrogate model of the 6D PES of adsorption and identified its minima, in which we detected the most stable structure (global minimum) and seven other stable structures (local minima).

We classified our stable structures into two classes, that is, Ox and Hy, with respect to the bonding species in the adsorbed camphor. The differences between the two classes were further categorized by the trends in the adsorption energies and the energy barriers of molecular motion. By analyzing the electronic structure of the stable adsorbates, we concluded that in the most stable structures (class Ox), camphor chemisorbs to the Cu surface via O bonding. Our results imply that class Ox structures are viable candidates for static camphor adsorbates observed in AFM experiments.

By combining machine learning with DFT, BOSS provides a novel method for a reliable structure identification via the surrogate model of the PES. With the complete PES, we obtain chemical insight into numerous materials properties (e.g., the stable adsorbate structures and their mobility) in one go, without prior presumptions about the material. Our approach eliminates the human bias present in conventional structure search, in which the optimal structures are commonly estimated using chemical intuition. Efficient and unbiased structure search methods, such as BOSS, facilitate the study of complex hybrid interface structures. The acquired knowledge can be applied in the precision engineering of interface structures in functional materials to optimize their advantageous properties.

Supporting Information

Supporting information features camphor geometry in global minimum conformer search, convergence of the 6D surrogate model, and coordinates of camphor in the predicted and relaxed stable structures.

Supporting Information File 1

Camphor global minimum conformer, convergence of the 6D model, and coordinates of camphor.

[<https://www.beilstein-journals.org/bjnano/content/supplementary/2190-4286-11-140-S1.pdf>]

Acknowledgements

The authors wish to acknowledge CSC – IT Center for Science, Finland, and the Aalto Science-IT project for generous computational resources. An award of computer time was provided by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.

Funding

This work has received funding from the Academy of Finland via the Artificial Intelligence for Microscopic Structure Search (AIMSS) project No. 316601 and the Flagship programme: Finnish Center for Artificial Intelligence FCAI, and from the Emil Aaltonen Foundation.

ORCID® iDs

Jari Järvi - <https://orcid.org/0000-0003-3727-5887>

Patrick Rinke - <https://orcid.org/0000-0003-1898-723X>

Milica Todorović - <https://orcid.org/0000-0003-0028-0105>

Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: <https://arxiv.org/abs/2002.05598>

References

1. Snaith, H. J. *Nat. Mater.* **2018**, *17*, 372–376. doi:10.1038/s41563-018-0071-z
2. Terzaki, K.; Kissamitaki, M.; Skarmoutsou, A.; Fotakis, C.; Charitidis, C. A.; Farsari, M.; Vamvakaki, M.; Chatzinikolaidou, M. *J. Biomed. Mater. Res., Part A* **2013**, *101A*, 2283–2294. doi:10.1002/jbm.a.34516
3. Contessotto, L.; Ghedini, E.; Pinna, F.; Signoretto, M.; Cerrato, G.; Crocellà, V. *Chem. – Eur. J.* **2009**, *15*, 12043–12049. doi:10.1002/chem.200900603
4. Farronato, M.; Bidermane, I.; Lüder, J.; Bouvet, M.; Vlad, A.; Jones, A.; Simbrunner, J.; Resel, R.; Brena, B.; Prévot, G.; Witkowski, N. *J. Phys. Chem. C* **2018**, *122*, 26480–26488. doi:10.1021/acs.jpcc.8b08462
5. Wang, L.; Bian, J.; Lu, L.; Liang, Z.; Zhang, D.; Yang, B.; Li, L.; Lu, G.; Yang, Y. *J. Mater. Chem. C* **2020**, *8*, 3527–3535. doi:10.1039/c9tc06739f
6. Kühnle, A. *Curr. Opin. Colloid Interface Sci.* **2009**, *14*, 157–168. doi:10.1016/j.cocis.2008.01.001
7. Kumar, A.; Banerjee, K.; Dvorak, M.; Schulz, F.; Harju, A.; Rinke, P.; Liljeroth, P. *ACS Nano* **2017**, *11*, 4960–4968. doi:10.1021/acsnano.7b01599
8. Tian, G.; Shen, Y.; He, B.; Yu, Z.; Song, F.; Lu, Y.; Wang, P.; Gao, Y.; Huang, H. *Surf. Sci.* **2017**, *665*, 89–95. doi:10.1016/j.susc.2017.08.008
9. Wold, D. J.; Haag, R.; Rampi, M. A.; Frisbie, C. D. *J. Phys. Chem. B* **2002**, *106*, 2813–2816. doi:10.1021/jp013476t
10. Mönig, H.; Amirjalayer, S.; Timmer, A.; Hu, Z.; Liu, L.; Díaz Arado, O.; Cnudde, M.; Strassert, C. A.; Ji, W.; Rohlfling, M.; Fuchs, H. *Nat. Nanotechnol.* **2018**, *13*, 371–375. doi:10.1038/s41565-018-0104-4

11. Zint, S.; Ebeling, D.; Schlöder, T.; Ahles, S.; Mollenhauer, D.; Wegner, H. A.; Schirmeisen, A. *ACS Nano* **2017**, *11*, 4183–4190. doi:10.1021/acsnano.7b01109
12. Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864–B871. doi:10.1103/physrev.136.b864
13. Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138. doi:10.1103/physrev.140.a1133
14. Goedecker, S. *J. Chem. Phys.* **2004**, *120*, 9911–9917. doi:10.1063/1.1724816
15. Li, F.; Liu, Y.; Wang, L.; Zhao, J.; Chen, Z. *Theor. Chem. Acc.* **2012**, *131*, 1163. doi:10.1007/s00214-012-1163-5
16. Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566. doi:10.1073/pnas.202427399
17. Obersteiner, V.; Scherbela, M.; Hörmann, L.; Wegner, D.; Hofmann, O. T. *Nano Lett.* **2017**, *17*, 4453–4460. doi:10.1021/acs.nanolett.7b01637
18. Packwood, D. M.; Han, P.; Hitosugi, T. *Nat. Commun.* **2017**, *8*, 14463. doi:10.1038/ncomms14463
19. Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning; Adaptive Computation and Machine Learning; The MIT Press: Cambridge, MA, USA, 2005.* doi:10.7551/mitpress/3206.001.0001
20. Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R. P.; de Freitas, N. *Proc. IEEE* **2016**, *104*, 148–175. doi:10.1109/jproc.2015.2494218
21. Garijo del Río, E.; Mortensen, J. J.; Jacobsen, K. W. *Phys. Rev. B* **2019**, *100*, 104103. doi:10.1103/physrevb.100.104103
22. Koistinen, O.-P.; Dagbjartsdóttir, F. B.; Ásgeirsson, V.; Vehtari, A.; Jónsson, H. *J. Chem. Phys.* **2017**, *147*, 152720. doi:10.1063/1.4986787
23. Seko, A.; Maekawa, T.; Tsuda, K.; Tanaka, I. *Phys. Rev. B* **2014**, *89*, 054303. doi:10.1103/physrevb.89.054303
24. Balachandran, P. V.; Xue, D.; Theiler, J.; Hogden, J.; Lookman, T. *Sci. Rep.* **2016**, *6*, 19660. doi:10.1038/srep19660
25. Chan, L.; Hutchison, G. R.; Morris, G. M. *J. Cheminf.* **2019**, *11*, 32. doi:10.1186/s13321-019-0354-7
26. Packwood, D. M.; Hitosugi, T. *Appl. Phys. Express* **2017**, *10*, 065502. doi:10.7567/apex.10.065502
27. Carr, S.; Garnett, R.; Lo, C. *BASC: Applying Bayesian Optimization to the Search for Global Minima on Potential Energy Surfaces. In Proceedings of Machine Learning Research, June 20–22, 2016; Balcan, M. F.; Weinberger, K. Q., Eds.; PMLR: New York, New York, USA, 2016; pp 898–907.*
28. Jørgensen, M. S.; Larsen, U. F.; Jacobsen, K. W.; Hammer, B. *J. Phys. Chem. A* **2018**, *122*, 1504–1509. doi:10.1021/acs.jpca.8b00160
29. Xue, D.; Balachandran, P. V.; Hogden, J.; Theiler, J.; Xue, D.; Lookman, T. *Nat. Commun.* **2016**, *7*, 11241. doi:10.1038/ncomms11241
30. Seko, A.; Togo, A.; Hayashi, H.; Tsuda, K.; Chaput, L.; Tanaka, I. *Phys. Rev. Lett.* **2015**, *115*, 205901. doi:10.1103/physrevlett.115.205901
31. Todorović, M.; Gutmann, M. U.; Corander, J.; Rinke, P. *npj Comput. Mater.* **2019**, *5*, 35. doi:10.1038/s41524-019-0175-2
32. Egger, A. T.; Hörmann, L.; Jeindl, A.; Scherbela, M.; Obersteiner, V.; Todorović, M.; Rinke, P.; Hofmann, O. T. *Adv. Sci.* **2020**, *7*, 2000992. doi:10.1002/advs.202000992
33. Fang, L.; Makkonen, E.; Todorović, M.; Rinke, P.; Chen, X. *arXiv* **2020**, No. 2006.15006.
34. BOSS code repository. <https://gitlab.com/cest-group/boss> (accessed March 13, 2020).
35. Alldritt, B.; Hapala, P.; Oinonen, N.; Urtev, F.; Krejci, O.; Federici Canova, F.; Kannala, J.; Schulz, F.; Liljeroth, P.; Foster, A. S. *Sci. Adv.* **2020**, *6*, eaay6913. doi:10.1126/sciadv.aay6913
36. Oganov, A. R.; Glass, C. W. *J. Chem. Phys.* **2006**, *124*, 244704. doi:10.1063/1.2210932
37. Gutmann, M. U.; Corander, J. *J. Mach. Learn. Res.* **2016**, *17*, 4256–4302.
38. Brochu, E.; Cora, V. M.; de Freitas, N. *arXiv* **2010**, No. 1012.2599.
39. Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. *SIAM J. Sci. Comput.* **1995**, *16*, 1190–1208. doi:10.1137/0916069
40. Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833–1840. doi:10.1063/1.1740588
41. Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. *Comput. Phys. Commun.* **2009**, *180*, 2175–2196. doi:10.1016/j.cpc.2009.06.022
42. Havu, V.; Blum, V.; Havu, P.; Scheffler, M. *J. Comput. Phys.* **2009**, *228*, 8367–8379. doi:10.1016/j.jcp.2009.08.008
43. Ren, X.; Rinke, P.; Blum, V.; Wierferink, J.; Tkatchenko, A.; Sanfilippo, A.; Reuter, K.; Scheffler, M. *New J. Phys.* **2012**, *14*, 053020. doi:10.1088/1367-2630/14/5/053020
44. Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868. doi:10.1103/physrevlett.77.3865
45. Ruiz, V. G.; Liu, W.; Tkatchenko, A. *Phys. Rev. B* **2016**, *93*, 035118. doi:10.1103/physrevb.93.035118
46. Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 073005. doi:10.1103/physrevlett.102.073005
47. Hofmann, O. T.; Atalla, V.; Moll, N.; Rinke, P.; Scheffler, M. *New J. Phys.* **2013**, *15*, 123028. doi:10.1088/1367-2630/15/12/123028
48. Ruiz, V. G.; Liu, W.; Zojer, E.; Scheffler, M.; Tkatchenko, A. *Phys. Rev. Lett.* **2012**, *108*, 146103. doi:10.1103/physrevlett.108.146103
49. van Lenthe, E.; Baerends, E. J.; Snijders, J. G. *J. Chem. Phys.* **1993**, *99*, 4597–4610. doi:10.1063/1.466059
50. Haas, P.; Tran, F.; Blaha, P. *Phys. Rev. B* **2009**, *79*, 085104. doi:10.1103/physrevb.79.085104
51. Liu, W.; Ruiz, V. G.; Zhang, G.-X.; Santra, B.; Ren, X.; Scheffler, M.; Tkatchenko, A. *New J. Phys.* **2013**, *15*, 053046. doi:10.1088/1367-2630/15/5/053046
52. Stroppa, A.; Termentzidis, K.; Paier, J.; Kresse, G.; Hafner, J. *Phys. Rev. B* **2007**, *76*, 195440. doi:10.1103/physrevb.76.195440
53. Järvi, J.; Alldritt, B.; Krejčí, O.; Todorović, M.; Liljeroth, P.; Rinke, P. *Res. Square* **2020**. doi:10.21203/rs.3.rs-50783/v1

License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>). Please note that the reuse, redistribution and reproduction in particular requires that the authors and source are credited.

The license is subject to the *Beilstein Journal of Nanotechnology* terms and conditions: (<https://www.beilstein-journals.org/bjnano>)

The definitive version of this article is the electronic one which can be found at:
<https://doi.org/10.3762/bjnano.11.140>