scientific reports

OPEN

Describing chain-like assembly of ethoxygroup-functionalized organic molecules on Au(111) using high-throughput simulations

Lokamani[1,2,3,8✉], Jeffrey Kelling[1,8], Robin Ohmann[3,4,8], Jörg Meyer[3], Tim Kühne[5], Gianaurelio Cuniberti[3], Jannic Wolf[6], Guido Juckeland[1], Thomas Huhn[6], Peter Zahn[2,8], Francesca Moresco[5] & Sibylle Gemming[5,7]

Due to the low corrugation of the Au(111) surface, 1,4-bis(phenylethynyl)-2,5-bis(ethoxy)benzene (PEEB) molecules can form quasi interlocked lateral patterns, which are observed in scanning tunneling microscopy experiments at low temperatures. We demonstrate a multi-dimensional clustering approach to quantify the anisotropic pair-wise interaction of molecules and explain these patterns. We perform high-throughput calculations to evaluate an energy function, which incorporates the adsorption energy of single PEEB molecules on the metal surface and the intermolecular interaction energy of a pair of PEEB molecules. The analysis of the energy function reveals, that, depending on coverage density, specific types of pattern are preferred which can potentially be exploited to form one-dimensional molecular wires on Au(111).

Molecules can form interlocked lateral patterns after being deposited on metal surfaces. The process of pattern formation may depend on various factors, in particular the competition between molecule-substrate and inter-molecular interaction. The formation of interlocked lateral patterns has been investigated recently using low temperature scanning tunneling microscopy (STM)[1], where laterally interlocked chains were grown epitaxially from inorganic ionic cyanides. In case of organic molecules with π-conjugated backbones adsorbed on Au(111) substrate, e.g. phenylenes and phenyleneethynylenes[2], the pattern-formation is mediated by a relatively weak van der Waals force and the direct interactions of the functional groups between the molecules. The interaction between organic molecules and inorganic substrate at the atomic level and interaction mediated pattern formation plays a crucial role in device characteristics[3]. It is essential to understand the underlying mechanism of such interactions in order to design and fabricate new molecules with enhanced functionality and stability for organic electronic devices[4].

In this article, we present the results of a combined theoretical and experimental study of elongated organic 1,4-bis(phenylethynyl)-2,5-bis(ethoxy)benzene (PEEB) molecules deposited on Au(111), which can also form quasi-interlocked lateral patterns on Au(111), as revealed through investigation by low temperature (STM). In order to investigate whether those patterns depend specifically on the molecular structure of PEEB, we scan the space of possible adsorption geometries by high-throughput calculations[5,6] using density-functional-based tight-binding plus (DFTB+)[7]. The present study follows the choice of high-symmetry adsorption sites for atoms and diatomic molecules in[8], but additionally employs the symmetry of the substantially larger molecules to define an irreducible wedge in configuration space and sample only that sector. We thus explore an energy landscape which describes the adsorption of single PEEB molecules on the Au(111) surface and the interaction within a molecular pair.
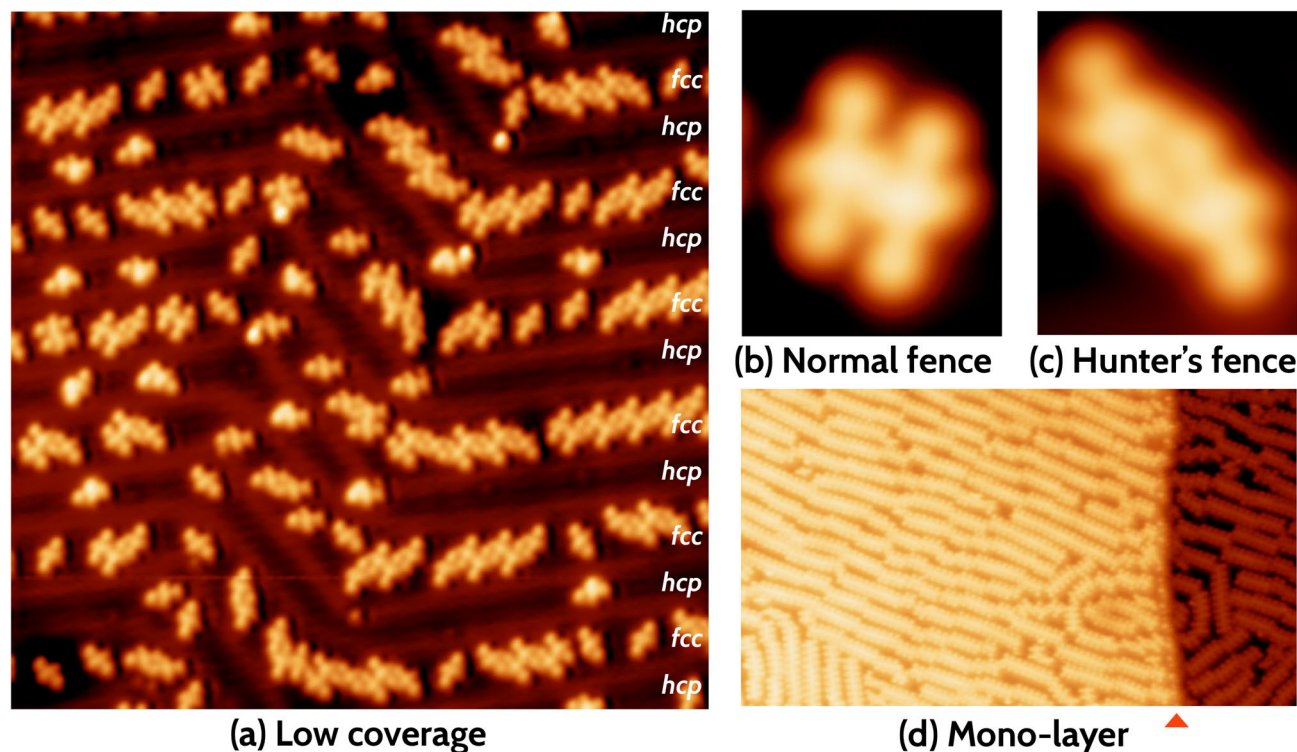
Figure 1. (**a**) STM images of PEEB adsorbed on Au(111) at low coverage (Image size: 40×40 nm^2, $V_{appl} = -0.5$ V, $I = 50$ pA). At low coverage PEEB molecules adsorb predominantly on fcc-regions. Two prominent clustering patterns are denoted as (**b**) *Normal fence* (image size: 2.8×3.5 nm^2, $V_{appl} = -0.1$ V, $I = 50$ pA) and (**c**) *Hunter's fence* (image size: 2.8×3.5 nm^2, $V_{appl} = -1$ V, $I = 50$ pA) (**d**) On increasing the coverage of the PEEB molecules, the *normal fence*-like clustering into long one-dimensional rows is observed only. The coverage of the one-dimensional rows extends until the step edges. Step edge shown using red triangle (image size: 70×40 nm^2, $V_{appl} = -1$ V, $I = 50$ pA). Note that, the atomic resolution of the substrate is not possible, while imaging adsorbed molecules because different tunneling parameters are employed. The exact adsorption position can therefore only be determined by density functional theory (DFT) calculations.

High throughput methods have recently gained increasing interest, because they allow the automated search for thermodynamically stable structural and magnetic phases[9–15] and also serve to determine average numerical parameters for classical and quantum mechanical simulation approaches[16–19]. Consequently, high-throughput methods to screen compounds for desirable physical properties of interest have become the major driving force behind numerous data driven materials prediction methods[20–24]. Rather than applying the aforementioned methods to screen a range of compounds, we employ a homogeneous sampling strategy to explore the configurational energy landscape of a single system with the objective of describing intricate steric and electronic interactions of molecular pairs to understand the building blocks of the experimentally observed pattern formation. We demonstrate a visual approach of decomposing the multiple dimensions of the energy landscape to track down all prominent pair formations.

Results

Experimental results. In the STM experiments, various kinds of molecular clustering are observed at different degrees of coverage for PEEB molecules on Au(111) at low temperatures. At a low degree of coverage, the PEEB molecules cluster predominantely into two *fence-like* structures: *normal fence* and *hunter's fence*, as shown in Fig. 1b,c, respectively. As the concentration of the molecules and with that the degree of coverage on the gold substrate is increased, only rows formed in a normal-fenced pattern are observed (see Fig. 1d). At low coverage, the adsorption occurs preferentially on the fcc-region of the reconstructed Au(111) surface.

Single PEEB molecule: adsorption on Au(111). A combined theoretical and experimental study of PEEB molecules deposited on Au(111) is presented in a recent publication[25]. The electronic and geometric structure of a single PEEB molecule on Au(111) surface is investigated using low-temperature STM. The density-functional-based tight-binding (DFTB)[7] simulations of the geometrical configuration, density of states (DOS), and charge-reordering and charge-transfer suggest, that PEEB molecules are physisorbed on Au(111). The work of adhesion equals 0.31 Jm^{-2}, implying a non-reactive wetting of Au(111) by PEEB molecules. Our analysis reveals not only a negligible change of geometrical structure upon adsorption but also negligible absolute charge transfer per atom of mere 0.003|e| to the substrate. The total charge transfer between the molecule and the sub-
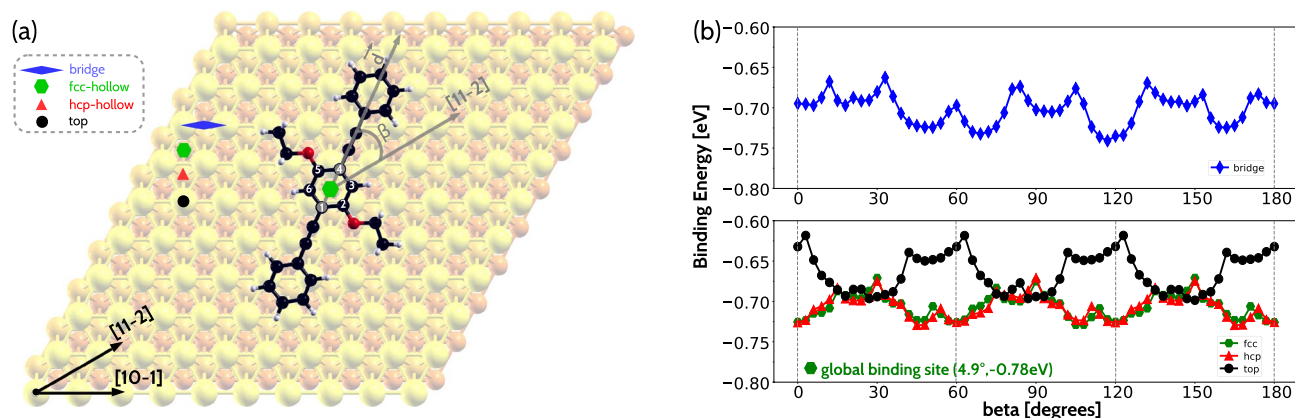
Figure 2. (**a**) Sampling shown for the central phenyl ring placed over an fcc-hollow site: β denotes the in-plane rotational angle of MB to the [11-2] direction. The vector d defining the MB-direction is spanned between the carbon atoms 1 and 4 of the central phenyl ring. β has the value 0°, if the molecular backbone is oriented parallel to the [11-2] direction. The bridge-, fcc-, hcp-, top-sites are shown using blue skewed diamond, green hexagon, red triangle and black circle, respectively. There is an atom directly beneath the third (second) layer below the fcc (hcp) hollow site, respectively. (**b**) Binding energies: The configurational spaces bridge-ms, fcc-ms, hcp-ms, and top-ms are denoted using blue, green, red, and black curves, respectively. fcc-ms, hcp-ms and top-ms curves reflect a three-to-six-fold rotational symmetry, whereas bridge-ms does not show any additional symmetry between 0° and 180°, as expected.

strate equals 0.14|e|, which excludes the possibility of intermolecular Coulomb interaction. Furthermore, the charge re-distribution for the molecule occurs predominantly along the side chains (SCs) and the outer phenyl ring (OPR), which is not perturbed significantly upon adsorption. An overall dipole moment is prohibited by the presence of an inversion center. Thus, a possible clustering of PEEB molecules would be driven by interactions between partial-charge centers.

High-throughput calculations: clustering. In order to explain the clustering of PEEB molecules on Au(111) we propose an energy function incorporating two concurrent yet independent processes: (a) a Surface Binding Energy, which describes the adsorption of single PEEB molecules on Au(111) on a chosen site and (b) a Planar Clustering Energy, which takes into consideration the intermolecular interactions solely between two isolated PEEB molecules. We define the sum of both energies as total energy E_{tot}.

We sample the configurational space in search for the most likely interlocking patterns in two steps: At first, we determine the energy landscape of the binding process of a single molecule on the surface. For that, we place an isolated PEEB molecule over Au(111) with the central phenyl ring (CPR) aligned over a high-symmetry site at the Au surface. The in-plane rotation angle β of the start configuration is choosen as 0°, when the molecular backbone (MB) is parallel to the [11-2] direction of the Au(111) surface, see Figure 2(a). Subsequently, β is increased to 180° in steps of 3°, while keeping the center of the CPR aligned to the chosen adsorption site. As the molecule exhibits an inversion center, only angles of at most 180° need to be scanned, thus the number of configurations considered are 61. Lower number of configurations are needed at sites with higher local symmetry of Au(111). The average distance between the carbon atoms and the topmost gold layer is set to 4 Å in the start configuration. The so-generated configurations are relaxed with the following geometrical constraints: a) the gold atoms of the substrate are kept fixed, b) the CPR is allowed to relax only in the z-direction and c) the remaining atoms of the molecule are relaxed without any constraints. This procedure is repeated for the inversion center in the middle of the central phenyl ring placed over an fcc-hollow, an hcp-hollow, a top and a bridge position. Four subgroups on fcc, hcp, top and bridge sites span the configurational space: fcc-ms, hcp-ms, top-ms and bridge-ms, where ms denotes molecular-substrate system. The constraint (b) ensures, that the final configurations are unique and the sampling of the configurational space is unbiased[26–29]. The choice of the supercell and the spurious interactions caused by the periodic boundary conditions are discussed in the supplemental information (SI).

The surface binding energies for the configurations are shown in Fig. 2b. As expected, the green, red and black curves corresponding to the binding energies of the fcc-, hcp- and top-configurational space resemble very closely a 6-fold symmetry. This is due to the fact, that a particular configuration and its counterpart rotated by 60° about the CPR in the fcc-ms, hcp-ms and top-ms configurations are geometrically equivalent and have essentially the same total energies. On the other hand, the blue curve corresponding to the binding energy of the bridge-configurational space does not show any rotational symmetry within the scanned angle range. This reflects the reduced local symmetry of the bridge position. Within the bridge-ms configurational space, a particular starting configuration and its counterpart rotated by 180° about the CPR are geometrically equivalent and have the same binding energy. The deviations in binding energies from the ideal symmetric case are visible at 90° as compared to at angles of 30° and 150° for the green, red and black curves. These deviations are caused by the difference in the strength of interaction between the molecule and its periodic images attributed to the alignment of the MB along the short and long diagonal of the supercell. These deviations depend on the form and dimension of the periodic supercell which induce a symmetry-break.
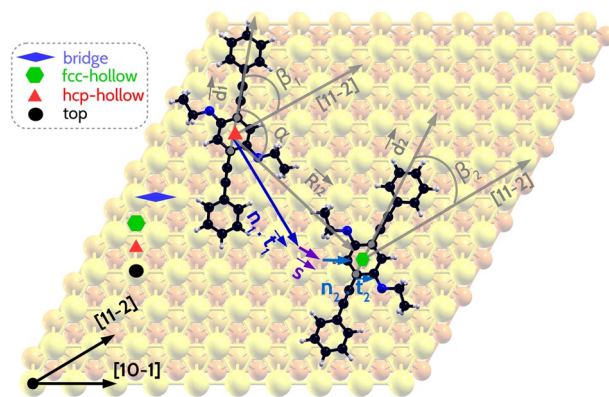
Figure 3. Formation of a molecular pair from the configurational space subgroup [hcp-m|fcc-m]: CPR of the first molecule is on a hcp-hollow site, of the second molecule on a fcc-hollow site. The in-plane rotation of the MB of the first molecule w.r.t [11-2] is $\beta_1 = 54°$, and of the second molecule $\beta_2 = 27°$. The total distance between the centers of mass of the molecules corresponds to $3\vec{t_1} + 1\vec{t_2} + \vec{s}$. The angle α, which is formed between the direction vector $\vec{d_1}$ of the first molecule and the vector $\vec{R_{12}}$ connecting the centers of mass of both the molecules amounts to 122.4°.

Next, in search for the most likely interlocking patterns as the second step, we utilize the single molecule configurations fcc-ms, hcp-ms, top-ms and bridge-ms to sample the configurational space of the clustering of molecule pairs to form interlocking patterns. The geometries of the molecules are extracted from the single molecule configurations excluding the gold surface to form pure molecular configurational space subgroups: fcc-m, hcp-m, top-m and bridge-m. A new set of configurations of molecular pairs is formed by performing the following steps: (1) a particular configuration is chosen from fcc-m, hcp-m, top-m or bridge-m, (2) another configuration is chosen from fcc-m, hcp-m, top-m or bridge-m, 3) the second configuration is translated w.r.t. the first molecule in units of the lattice vectors $\vec{t_1} = [1, \bar{1}, 0]$ and $\vec{t_2} = [1, 0, \bar{1}]$. Steps 1 and 2 ensure that all pairs of in-plane rotations of the MBs relative to the [11-2] surface direction are taken into account. Translations in units of the lattice vectors ensure that the second molecule preserves the geometrical invariance w.r.t. the chosen adsorption site for the inversion center of the PEEB molecule. These steps are repeated for all possible pairs of the pure molecular configurational subgroups: fcc-m, hcp-m, top-m and bridge-m. Clusters generated from identical configurational subgroups do not pose a binding site mismatch of the CPR, since both of the CPRs are centered on identical fcc-, hcp-hollow, top or bridge sites, respectively. In order to account for a binding site mismatch of the CPRs, which arises while generating a cluster of molecular pairs from different configurational subgroups e.g. [hcp-m|fcc-m], an additional shift vector $\vec{s}$ is applied to the translation of the second molecule. The shift vector in this case would be $\vec{s} = \frac{1}{3}[\vec{t_1} + \vec{t_2}]$. The procedure is schematically shown in Fig. 3.

This procedure is applied for all possible unique pair combinations (in total 10). The sampled configuration space for molecule pairs (MP-CS) is spanned by (a) a shift vector $\vec{R_{12}}$, which is spanned between the centers of mass of the molecules, (b) the difference between the angle of in-plane rotation of the individual molecules w.r.t. the [11-2] direction $\Delta\beta = \beta_1 - \beta_2$, and (c) the angle α, which is defined as the angle between the following two vectors: (1) the directional vector $\vec{d_1}$ of the first molecule and (2) the vector spanned between the center of mass of the first molecule and the center of mass of the second molecule $\vec{R_{12}}$, shown in Fig. 3. The choice of the supercell is discussed in the SI (see Fig. S2).

Analysis of molecule pair configurations. Figure 4a,b show the total energy over distance R_{12} for all configurations. This reveals three local minima which are enclosed by the intervals numbered 1, 2 and 3, indicating corresponding groups of configurations: group 1 ($8\text{Å} < R_{12} < 11\text{Å}$), group 2 ($11\text{Å} < R_{12} < 14\text{Å}$) and group 3 ($14\text{Å} < R_{12} < 16\text{Å}$). Panel (a), colored according to α, shows, that each of these three peaks is dominated by configurations with distinct relative alignments at angles $\alpha \approx 80°$ for group 1, 30° for group 2 and 150° for group 3 w.r.t. the connection vector $\vec{R_{12}}$. The inset of panel (a) visualizes those three major orientations and indicates the presence of further weakly preferable configurations with $\alpha \approx 0°$ and 130°, occurring for distances of 16Å and beyond. Panel (b) indicates the relative in-plane rotation $\Delta\beta$ of the two molecules depicted using color. The inset indicates that the relative angles $\Delta\beta = 0°$ and 180° (parallel configurations) have a clear prevalence over the other angles. Given the elongation of the MB and its high aspect ratio of about 4:1, this finding reflects the geometrical constraints when adsorbing PEEB molecules on surfaces with 1D-dominated reconstructions such as the herringbone structure of Au(111).

Figure 4c-h decompose the identified groups further by presenting the energy over α in panels (c) to (e), and over $\Delta\beta$ in panels (f) to (h). For each distance group separately, we defined intervals around the pronounced $E_{tot}(\alpha)$-peaks and $E_{tot}(\Delta\beta)$-peaks for our interval outer product ansatz. The resulting minima of all interval products are labelled and marked. The two lowest minima out of each $E_{tot}(\alpha)$-peak are highlighted by bold font, the corresponding molecular-pair geometries are displayed in the supplemental material.

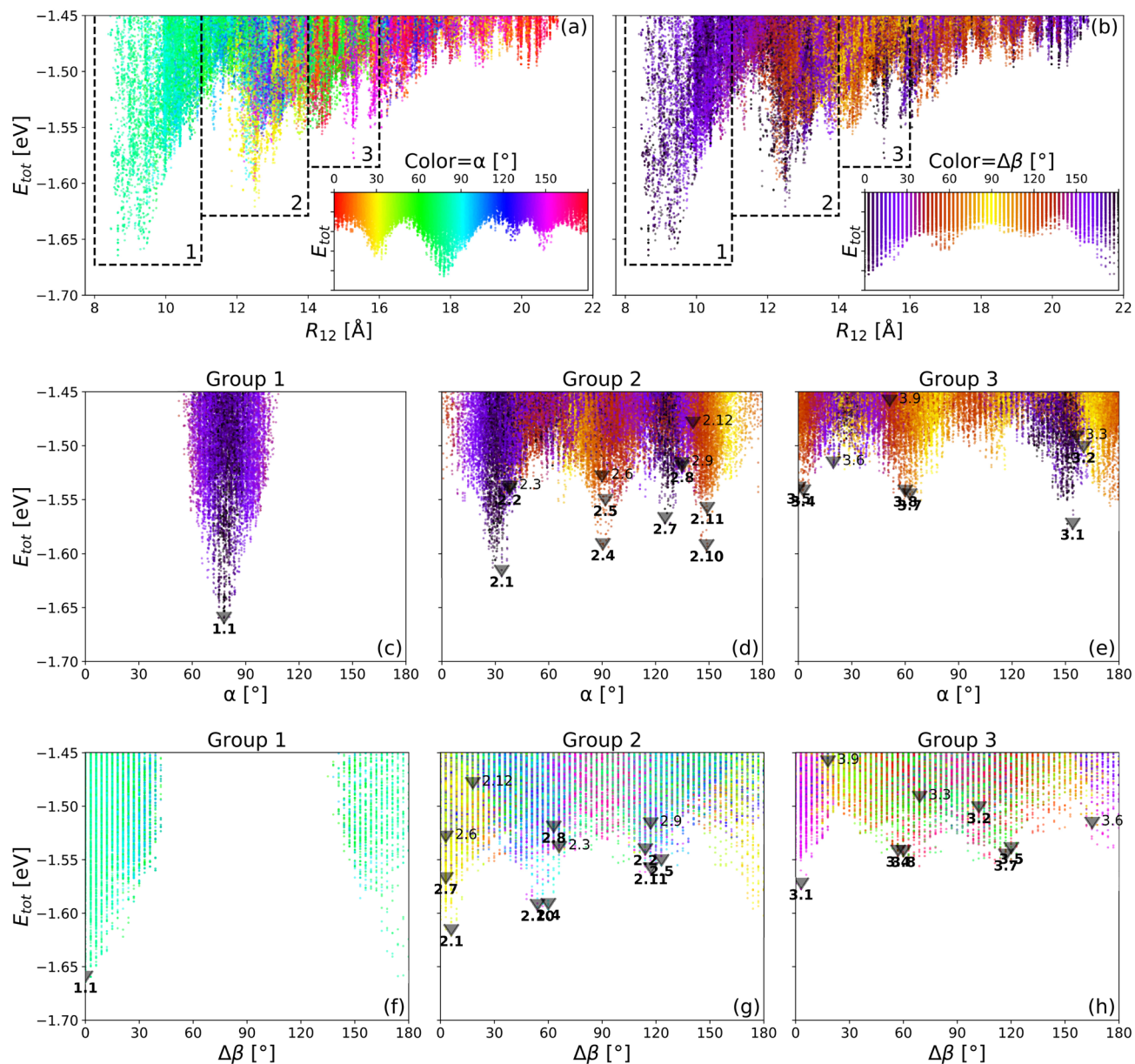**Figure 4.** (**a,b**): Total energy E_{tot} for pair configurations over distance R_{12}. The color maps show the position of one molecule relative to the other described by the angle α enclosed by one molecular backbone and the connection line, and the relative orientation angle $\Delta\beta$ between the molecules, respectively. The insets show the colorscales for the angles as $E_{tot}(\alpha)$ and $E_{tot}(\Delta\beta)$ plots. The insets reflect intrinsic symmetries of the system which are invariance against a rotation by $180°$ in α, panel (a), and mirror symmetry around $\Delta\beta = 0°$ and $90°$, panel (**b**). The marked y-values in the insets are the same as in the main figure. Dashed frames mark the regions around the main peaks selected as distance groups for further analysis. (**c-e**): Total energy E_{tot} for pair configurations over α. Each panel shows configurations from one distance group. The labeled triangles indicate the positions of sampled configurations for each visible peak. The colors correspond to the relative orientation of the molecules by angle $\Delta\beta$, see panel (**b**) for the color key. (**f-h**): Total energy E_{tot} for pair configurations over the relative orientation angle $\Delta\beta$ of the molecular backbones. Each panel shows configurations from one distance group only. The labeled triangles indicate the positions of sampled configurations for each visible peak. The colors correspond to the angle α, see panel (**a**) for the color key.

In Group 1, there is only one symmetrically unique minimum state, labelled 1.1, representing a configuration in which the molecules are parallel and only slightly offset ($\alpha \approx 80°$) at a distance of about 9Å. We denote this configuration *normal fence*. A prototype is shown in Fig. 5a.

In Group 2, corresponding to larger distances of about 12.5Å between the centers, the *normal fence* configuration is no longer favorable. The minimum energy configuration, although still a parallel one, is strongly offset at $\alpha \approx 33.7°$ as indicated by the label 2.1. This configuration is shown as prototype in Fig. 5b and we denote it as *hunter's fence A*. The runner-up configurations 2.2 and 2.3 could be considered marginal as they are close to
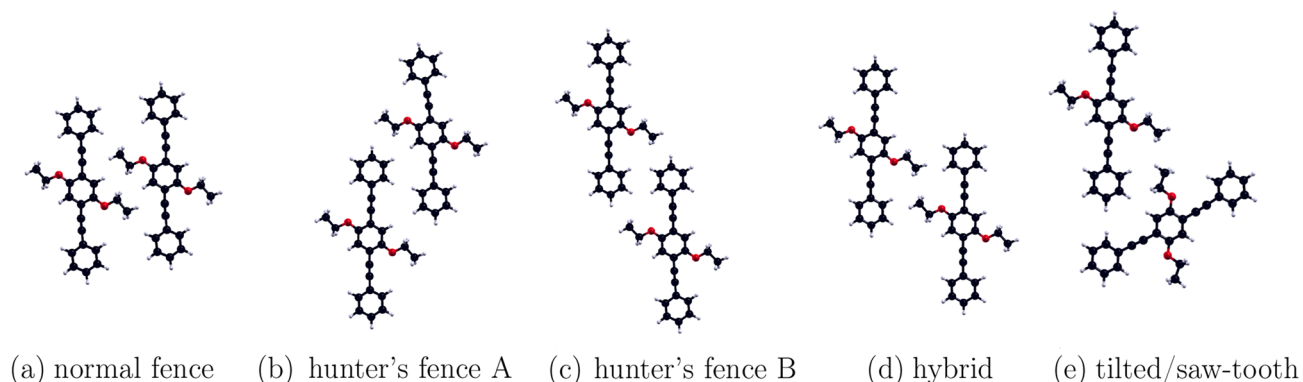
(a) normal fence   (b) hunter's fence A   (c) hunter's fence B   (d) hybrid   (e) tilted/saw-tooth

Figure 5. Prototypes for energetically most favorable pair configurations (**a**) *normal fence* 1.1 (**b**) *hunter's fence A* 2.1 (**c**) *hunter's fence B* 3.1 (**d**) *hybrid* 2.7 and (**e**) *saw-tooth* 2.4.

the energy (≈ -1.55 eV) above which no peaks were considered. Another much less stable parallel configuration labelled 2.7 belongs to the peak with a principal angle α of about 125.4°. Because this angle is between that of the *normal* and *hunter's fence* configurations we label it *hybrid*. A prototype of this is shown in Fig. 5d. Figure 4g shows subsequently energetically favorable non-parallel ($\Delta\beta \neq 0°, 180°$) configurations, representing energy minima in the peaks marked 2.4 and 2.10, the runner-up configurations here, belong to a symmetrically equivalent peak at $\Delta\beta \approx 60°, 120°$. The prototype 2.4 of this configuration, which we denote *saw-tooth*, is shown in Fig. 5e. The remaining second runner-ups are not of interest because their total energy is clearly above the base of their respective peaks.

In Group 3, we restrict to a single relevant peak with the prototype 3.1. It is similar to *hunter's fence A* (see Fig. 5b), except that with an offset of 150° in α, the phenyl heads dock at the acute angle between arms and backbone of the other molecule, necessitating the larger distance. Due to this similarity, we denote this configuration *hunter's fence B* shown in Fig. 5c. These two configurations are complementary in that, if the molecules were to densely cover a flat surface with *hunter's fence A* chains, the perpendicular coordination would have to be *hunter's fence B*.

All marked/selected configurations in Fig. 4 are shown explicitly in SI section II.

Discussion: clustering models. The STM images reveal that the molecules adsorb preferentially on the fcc-domains of the herringbone reconstruction of the Au(111) surface. The preference of organic molecules for the fcc-domains on Au(111) has previously been reported in the literature[30]. Although no specific orientation of single molecules adsorbed on Au(111) is discernible, the molecules group together in two different configurations to form molecular clusters and chains (see Fig. 1). In the first configuration (Fig. 1b), the molecules group together in such a way that the side chains lie next to each other to form *normal-fence*-like structures. In the second configuration (Fig. 1c), the molecules group together with the side chains and the outer phenyl rings lying next to each other to form *hunter's-fence*-like structures. At higher coverage the monolayer consists of the *normal-fence*-like structures only.

The analysis of the energy function supports the correlation of the adsorption pattern with the adsorbate density. If the distance between the central phenyl rings of adjacent molecules is between 8 Å and 11 Å, only *normal-fence*-like structures with mostly parallel molecules are preferred, as evident from the clustering analysis of group 1 (see Section III). This is in agreement with the STM observations of monolayers of PEEB molecules on Au(111), where the distance between the central phenyl rings of the molecules is minimum. The monolayers are comprised of rows of *normal-fence*-like structures as shown in Fig. 1d. For lower coverage of PEEB molecules on Au(111) and as the distance between the central phenyl rings of molecules increases, the *hunter's-fence*-like and *hybrid* structures occur, in addition to the *normal-fence*-like structures. These characteristics of the groups 2 and 3 are in accordance with the STM observations of *normal-fence*-like and *hunter's-fence*-like clusters at lower concentration as shown in Fig. 1a. Our calculations predict two forms of *hunter's-fence*-like structures: *hunter's fence A* and *hunter's fence B*. Furthermore, our calculations suggest that *hunter's fence A* is energetically more favorable than *hunter's fence B*. Given the sparcity of experimental statistical data on the orientation and proximity of individual molecules to surrounding molecules, a similar distinction of *hunter's-fence*-like structures in STM images is not viable.

Note that, Au adatoms can play an important role towards the formation of metalorganic structures on surface. Although, in most cases such Au adatoms are not visible in STM images[31–33], their presence is however demonstrated by DFT calculations and simulated imaging. In the present case, the theory reproduces the cluster formation by interactions between the centers of partial charge centers on the PEEB molecule without including and Au adatoms.

In summary, the *high-throughput* calculations and the evaluation of the one- and two-particle terms of the *energy function* E_{tot} attributed to the adsorption and clustering processes, elucidate the different principles involved in pattern formation of PEEB on Au(111) at various degrees of coverage.

Methods

Experimental setup. PEEB molecules were studied on the Au(111) surface with a low temperature STM at 5 K under ultrahigh vacuum conditions ($< 10^{-10}$ mbar). The gold single crystal surface was cleaned by repeated cycles of Ne ion sputtering and subsequent annealing at 720 K. The molecules were deposited from a Knudsen cell at 420 K onto the Au(111) substrate which was kept at room temperature for about 10min. After the deposition, which takes place in a dedicated preparation chamber, the sample is cooled down with liquid nitrogen to about 85 K and then transferred into the STM chamber, without disrupting the vaccum conditions. The time needed to cool down the sample to 5 K is of the order of 10 h in the present setup. A bias voltage was applied to the sample, while the tip was grounded.

Electronic structure methods. We performed self-consistent density-functional-based tight-binding simulations for geometric and electronic structural properties as implemented in the program package DFTB+[7]. The parameter set "auorg-1-1" has been utilized in all our calculations[34], which is an extension of the "mio-1-1"[35] parameter set to include gold. The "mio-1-1" set has been developed for organic molecules including O, N, C, H, and S atoms and works well for conformational energies and geometries of H-bonded systems[36,37]. The maximum angular momenta included in our calculations are: H:s, C:p, N:p, O:p and Au:d. In case of an isolated molecule and the geometries with gold substrate, the unit cell was spanned using following lattice vectors: $a_1 = [28.85\text{Å}, 0, 0]$, $a_2 = [14.43\text{Å}, 24.985\text{Å}, 0]$ and $a_3 = [0, 0, 25\text{Å}]$. The Au(111) substrate was modelled using three layers ($10 \times 10 \times 3$)[3,31], which were kept fixed (see Fig. S1 in the supplements). In case of pairs of molecules, the unit cell was doubled in both the a_1 and the a_2 directions. Periodic boundary conditions were used for all calculations. The relaxations were performed until the maximum force components reduced to 0.0001 eV/Å. All structural relaxation calculations were performed using the conjugate gradient algorithm including a universal force field extension for dispersion corrections as implemented in DFTB+[38].

Determining minimum configurations. We explore the energy landscape formed by adsorbed molecule pairs employing an intuitive graphical approach, which non the less guarantees, that all significant local minima, i.e. relevant pair configurations, are located regardless of the dimensionality of the problem. An absorbed pair can be characterized by three quantities, which we will define later: distance, relative position and relative orientation of the two molecules, giving rise to a hypersurface. Plotting total energy over each quantity separately will, in general, only reveal a subset of all significant local minima as pronounced peaks. Intervals enclosing a single peak each constitute regions of interest in each dimension. These intervals are defined to enclose the major part of the protrusion of the inverted peaks that are distinctly visible. Taking an outer product of the sets of intervals of all dimensions produces a super set of product intervals which encloses all significant local minima. The minimum energy configuration in each of these product intervals is either a local minimum configuration represented or hidden by one of the initially selected peaks, or an irrelevant higher-energy configuration. In order to find and tag all locally stable configurations, one can discard all product intervals the minimum energy configuration of which is above a baseline energy and above which no peaks were selected at the beginning. In order to better visualize how configurations relate to each other and better resolve peaks by dimension, we instead split our data into the intervals defined over distance first and marked product interval minima in plots over the remaining two dimensions in Sect. 2.4.

References

1. Lee, Y. *et al.* Universal oriented van der Waals epitaxy of 1d cyanide chains on hexagonal 2d crystals. *Adv. Sci.* **7**, 1900757. https://doi.org/10.1002/advs.201900757 (2020).
2. Riede, M., Mueller, T., Tress, W., Schueppel, R. & Leo, K. Small-molecule solar cells—status and perspectives. *Nanotechnology* **19**, 424001. https://doi.org/10.1088/0957-4484/19/42/424001 (2008).
3. Meyer, J. et al. Molecules for organic electronics studied one by one. *Phys. Chem. Chem. Phys.* **13**, 14421–14426. https://doi.org/10.1039/C1CP20999J (2011).
4. Chu, Y., Qian, C., Chahal, P. & Cao, C. Printed diodes: Materials processing, fabrication, and applications. *Adv. Sci.* **6**, 1801653. https://doi.org/10.1002/advs.201801653 (2019).
5. Zhu, Y. *et al.* Deepscreen: An accurate, rapid, and anti-interference screening approach for nanoformulated medication by deep learning. *Adv. Sci.* **5**, 1800909, https://doi.org/10.1002/advs.201800909 (2018). https://onlinelibrary.wiley.com/doi/pdf/10.1002/advs.201800909.
6. Himanen, L., Geurts, A., Foster, A. S. & Rinke, P. Data-driven materials science: Status, challenges, and perspectives. *Adv. Sci.* **6**, 1900808, https://doi.org/10.1002/advs.201900808 (2019). https://onlinelibrary.wiley.com/doi/pdf/10.1002/advs.201900808.
7. Aradi, B., Hourahine, B. & Frauenheim, T. Dftb+, a sparse matrix-based implementation of the dftb method. *J. Phys. Chem. A* **111**, 5678–5684. https://doi.org/10.1021/jp070186p (2007).
8. Montoya, J. H. & Persson, K. A. A high-throughput framework for determining adsorption energies on solid surfaces. *NPJ Comput. Mater.* **3**, 14. https://doi.org/10.1038/s41524-017-0017-z (2017).
9. Blatov, V. A. *et al.* High-throughput systematic topological generation of low-energy carbon allotropes. *NPJ Comput. Mater.* **7**, 15. https://doi.org/10.1038/s41524-021-00491-y (2021).
10. Horton, M. K., Montoya, J. H., Liu, M. & Persson, K. A. High-throughput prediction of the ground-state collinear magnetic order of inorganic materials using density functional theory. *NPJ Comput. Mater.* **5**, 64, https://doi.org/10.1038/s41524-019-0199-7 (2019).
11. Lenz, M.-O. *et al.* Parametrically constrained geometry relaxations for high-throughput materials science. *NPJ Comput. Mater.* **5**, 123, https://doi.org/10.1038/s41524-019-0254-4 (2019).

12. Zhang, Y. *et al.* High-throughput 3d reconstruction of stochastic heterogeneous microstructures in energy storage materials. *NPJ Comput. Mater.* **5**, 11, https://doi.org/10.1038/s41524-019-0149-4 (2019).
13. Kabiraj, A., Kumar, M. & Mahapatra, S. High-throughput discovery of high curie point two-dimensional ferromagnetic materials. *NPJ Comput. Mater.* **6**, 35, https://doi.org/10.1038/s41524-020-0300-2 (2020).
14. Torelli, D., Moustafa, H., Jacobsen, K. W. & Olsen, T. High-throughput computational screening for two-dimensional magnetic materials based on experimental databases of three-dimensional compounds. *NPJ Comput. Mater.* **6**, 158, https://doi.org/10.1038/s41524-020-00428-x (2020).
15. Sauceda, D. *et al.* High-throughput reaction engineering to assess the oxidation stability of max phases. *NPJ Comput. Mater.* **7**, 6, https://doi.org/10.1038/s41524-020-00464-7 (2021).
16. Choudhary, K. *et al.* High-throughput density functional perturbation theory and machine learning predictions of infrared, piezoelectric, and dielectric responses. *NPJ Comput. Mater.* **6**, 64, https://doi.org/10.1038/s41524-020-0337-2 (2020).
17. Wang, Z. *et al.* Machine learning method for tight-binding hamiltonian parameterization from ab-initio band structure. *NPJ Comput. Mater.* **7**, 11, https://doi.org/10.1038/s41524-020-00490-5 (2021).
18. Vitale, V. *et al. Automated high-throughput wannierisation. NPJ Comput. Mater.* **6**, 66. https://doi.org/10.1038/s41524-020-0312-y (2020).
19. Yu, M., Yang, S., Wu, C. & Marom, N. Machine learning the hubbard u parameter in dft+u using bayesian optimization. *NPJ Comput. Mater.* **6**, 180, https://doi.org/10.1038/s41524-020-00446-9 (2020).
20. Zunger, A., Wei, S.-H., Ferreira, L. G. & Bernard, J. E. Special quasirandom structures. *Phys. Rev. Lett.* **65**, 353–356. https://doi.org/10.1103/PhysRevLett.65.353 (1990).
21. Oses, C., Toher, C. & Curtarolo, S. High-entropy ceramics. *Nat. Rev. Mater.* **5**, 295–309. https://doi.org/10.1038/s41578-019-0170-8 (2020).
22. Addicoat, M. A., Coupry, D. E. & Heine, T. Autografs: Automatic topological generator for framework structures. *J. Phys. Chem. A* **118**, 9607–9614. https://doi.org/10.1021/jp507643v (2014).
23. Yang, S. *et al.* Ogre: A python package for molecular crystal surface generation with applications to surface energy and crystal habit prediction. *J. Chem. Phys.* **152**, 244122. https://doi.org/10.1063/5.0010615 (2020).
24. Curtarolo, S. et al. Aflow: An automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226. https://doi.org/10.1016/j.commatsci.2012.02.005 (2012).
25. Lokamani, *et al.* A combined experimental and theoretical study of 1,4-bis(phenylethynyl)-2,5-bis(ethoxy)benzene adsorption on Au(111). *Surf. Sci.* **121877**, 121877. https://doi.org/10.1016/j.susc.2021.121877 (2021).
26. Ferrenberg, A. M., Landau, D. P. & Binder, K. Statistical and systematic errors in Monte Carlo sampling. *J. Stat. Phys.* **63**, 867–882. https://doi.org/10.1007/BF01029988 (1991).
27. Landau, D. P. Finite-size behavior of the ising square lattice. *Phys. Rev. B* **13**, 2997–3011. https://doi.org/10.1103/PhysRevB.13.2997 (1976).
28. Tokdar, S. T. & Kass, R. E. Importance sampling: A review. *WIREs Comput. Stat.* **2**, 54–60, https://doi.org/10.1002/wics.56 (2010).
29. Kunze, T., Gemming, S., Numazawa, S. & Schreiber, M. Low-temperature modeling for degenerate and frustrated heisenberg systems with anisotropy. *Comput. Phys. Commun.* **181**, 806–812. https://doi.org/10.1016/j.cpc.2009.12.013 (2010).
30. Böhringer, M. *et al.* Two-dimensional self-assembly of supramolecular clusters and chains. *Phys. Rev. Lett.* **83**, 324–327. https://doi.org/10.1103/PhysRevLett.83.324 (1999).
31. Meyer, J. *et al.* Tuning the formation of discrete coordination nanostructures. *Chem. Commun.* **51**, 12621–12624. https://doi.org/10.1039/C5CC02723C (2015).
32. Robles, R. *et al.* Supramolecular chemistry based on 4-acetylbiphenyl on au(111). *Phys. Chem. Chem. Phys.* **22**, 15208–15213. https://doi.org/10.1039/D0CP01657H (2020).
33. Kühne, T. *et al.* Stm induced manipulation of azulene-based molecules and nanostructures: The role of the dipole moment. *Nanoscale* **12**, 24471–24476. https://doi.org/10.1039/D0NR06809H (2020).
34. Fihey, A. *et al.* Scc-dftb parameters for simulating hybrid gold-thiolates compounds. *J. Comput. Chem.* **36**, 2075–2087. https://doi.org/10.1002/jcc.24046 (2015).
35. Richard, R. M. & Herbert, J. M. Time-dependent density-functional description of the 1la state in polycyclic aromatic hydrocarbons: Charge-transfer character in disguise?. *J. Chem. Theory Comput.* **7**, 1296–1306. https://doi.org/10.1021/ct100607w (2011).
36. Elstner, M. *et al.* Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **58**, 7260–7268. https://doi.org/10.1103/PhysRevB.58.7260 (1998).
37. Niehaus, T., Elstner, M., Frauenheim, T. & Suhai, S. Application of an approximate density-functional method to sulfur containing compounds. *J. Mol. Struct. Theochem.* **541**, 185–194. https://doi.org/10.1016/S0166-1280(00)00762-4 (2001).
38. Rappe, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A. & Skiff, W. M. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10024–10035. https://doi.org/10.1021/ja00051a040 (1992).

Acknowledgements

Author contributions

M.L. designed the model and the computational framework. M.L., J.K., P.Z., G.J. and S.G. analysed the data and developed the clustering model. R.O., J.M., T.K., G.C. and F.M. conceived the STM experiments. J.W.and T.H. fabricated the PEEB molecules. M.L. wrote the manuscript with input from all authors. All authors reviewed the manuscript.

Funding

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/s41598-021-93724-5.

Correspondence and requests for materials should be addressed to L.

Reprints and permissions information is available at www.nature.com/reprints.

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