Supporting Information for

Structure Prediction for Surface-Induced Phases of Organic Monolayers: Overcoming the Combinatorial Bottleneck

Veronika Obersteiner,[†] Michael Scherbela,[†] Lukas Hörmann,[†] Daniel Wegner,[§] and Oliver T. Hofmann^{†,*}

[†] Institute of Solid State Physics, NAWI Graz, Graz University of Technology, Petersgasse 16, 8010 Graz, Austria

[§] Institute for Molecules and Materials, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands

Corresponding Authors

*o.hofmann@tugraz.at

The Supporting Information file includes:

I.	Supporting Methods	S2
١١.	Supporting Figures 1-12	S 5

I. Supporting Methods

Details to the DFT Methodology

As described in the main text DFT calculations were performed applying the FHI-aims code. For the calculation of the different polymorphs that partly lie very close in energy a huge accuracy of the energies is required. Therefore, we thoroughly performed the following convergence tests. For both, the unit cell for the local adsorption geometries (a $6\sqrt{3}x6\sqrt{3}$ R30 Au(111) surface unit cell) as well as for the polymorphs (the unit cell measured from the experimental STM with an epitaxy matrix of $\begin{pmatrix} 5 & 1 \\ -1 & 6 \end{pmatrix}$, we converged the adsorption energy per molecule with respect to the basis set, the k-point grid, the vacuum layer as well as the number of substrate layers within 0.01 eV per molecule. (For the local adsorption geometries we also converged the size of the unit cell to avoid spurious interactions between individual molecules.) The convergence criteria for the self-consistent cycles were set to 10^{-2} e/Å³ for the charge density and 10^{-5} eV for the total energy. We included relativistic effects within the ZORA approximation due to the presence of the gold substrate. The repeated slab approach with a vacuum of 60 Å was applied together with a dipole correction perpendicular to the surface.

Geometry optimizations were performed at the PBE level using the trust radius method enhanced version of the Broyden-Fletcher-Goldfarb-Shanno optimization algorithm with a tolerance of 10^{-2} eV/Å, until the remaining forces were less than 10^{-1} eV/Å. For the evaluation of the local adsorption geometries, we started from 48 different initial guesses varying the orientation of the molecule as well as the position with respect to the surface. For each orientation, i.e. flat lying TCNE, upright standing with the central C=C bond parallel to the surface and equivalently normal to the surface, we place the molecule on the four symmetry inequivalent adsorption sites (on top, fcc hollow, hcp hollow and bridge position) and moreover rotate the whole molecule around the z-direction, i.e. perpendicular to the substrate, by 0°, 15° , 30° and 45° . During the optimization the uppermost gold layer is relaxed together with the molecule. The 48 starting points converged into 9 distinct local adsorption geometries which are shown in Figure S1.

Details to the Monte Carlo Procedure

As mentioned in the main text, to explore the PES, we implemented an iterative Monte Carlo procedure. Each time a random neighbor is chosen from the neighbor list, this suggested polymorph is accepted or rejected on basis of its energy difference to the last accepted polymorph according to the Metropolis-Hastings scheme:

$$p_{acc} = min \left[1, exp \left(\frac{-(E_{current} - E_{last})}{k_B T} \right) \right) \right]$$
(1)

Hence, the suggested step is always accepted if its energy is lower than the last accepted one, $E_{current} < E_{last}$. For $E_{current} > E_{last}$ the new structure is only accepted if $p_{acc} > r$ where r represents a random number between 0 and 1. The temperature T is chosen such that the acceptance probability in the Metropolis criterion is high enough to allow the system to overcome barriers between local minima on the PES. Hence, it is possible to move uphill as well as downhill along the PES, providing a chance to escape from a local minimum energy in favor of finding a better, more global one. The starting temperature is set to 300 K and is adjusted during the simulation, similar to the flooding technique, *i.e.* it is decreased (increased) by 100 K in case of acceptation (rejection). While the temperature decrease upon acceptation helps to remain in energetically favorable regions around local minima, an increase in temperature upon rejection allows to jump out of a funnel. Applying this procedure, the overall acceptance rate adjusts to around 0.5, which means that half of all polymorphs are accepted, which is desired in traditional Basin Hopping.

Assumptions and their Limitations of SAMPLE

Every structure search algorithm needs to rely on a method to obtain total energies. For interfaces, these are band-structure calculations with periodic boundary conditions. Since these only accept one unique set of lattice vectors, they are limited to **commensurate** interfaces. (Calculating incommensurable interfaces, which are of course very common, is frequently done via compressing and stretching of one component. While this might have only minor implications on many observable properties, already small amounts of strain will cause large changes in the total energies. Hence, this is not a tractable solution for structure search approaches). In turn, because structure search is restricted to commensurate interfaces anyways, we can exploit additional properties that this kind of systems exhibit (see below): Most importantly, we rely on the fact that in commensurate structures, each molecule can be assigned a specific adsorption site on the surface. This is then exploited in the present approach in order to discretize the configuration space in an advantageous way. We do emphasize, once again, that this is not a limitation of our present approach: It is, rather, a limitation for structure search with periodic boundary conditions as a whole.

One assumption that we invoke is that said adsorption site is essentially independent of the molecular coverage: In other words, if an isolated molecule likes to adsorb in a specific position, the adsorption position in the tightly packed layer will be similar. Importantly, since the discretization procedure only provides starting points for the subsequent geometry

optimization, how "similar" these positions indeed are will only affect the efficiency of our approach, but not its overall validity. Very similar positions will be found in few geometry optimization steps, more dissimilar adsorption sites in the loosely and tightly packed case will require some steps more.

Once case where this approach is likely to go qualitatively awry is when in the tightly packed layer, new stable adsorption sites are generated in addition (rather than as a replacement) to the ones that exist for the isolated molecules. This situation is conceivable for strong molecular-molecular interactions; however, such systems are unlikely to form commensurate structures in the first place, and hence excluded from first-principles structure search anyways (see above).

Similarly, we assume that the adsorbate-adsorbate interaction does not cause large changes in the internal geometry of the adsorbate. This assumption is clearly well-placed in the case of (mostly rigid) conjugated organic molecules; it might fare worse if the molecules contain very flexible groups, such as aliphatic side chains.

In order to setup the configuration space, our algorithm requires four inputs: The size of the unit cell, the shape of the unit cell, the number of molecules (i.e., the coverage), and the minimum distance between atoms of different molecules. It is easily possible to setup systems for various sizes and shapes and compare the results for all of these, i.e. the first three parameters can be readily converged. We would also like to stress that because the interface must be commensurate, only a finite number of unit cell shapes is possible for a given cell size. Only the fourth parameter, the distance between molecules, is indeed a free parameter, that ought to be conservatively chosen (e.g. in the order of the covalent radii of the atoms).

II. Supporting Figures



Figure S1. Structure of the nine stable local adsorption geometries of TCNE/Au(111) ordered with respect to energy, i.e. structure G1 is the most stable local adsorption geometry (compare Figure S2). They differ in the orientation with respect to the surface, i.e. flat lying or upright standing, as well as in the particular binding site. As mentioned in the main text, we categorize them into three different groups: the flat lying adsorption geometries, the upright standing ones with the central C=C bond oriented parallel to the surface and equivalently normal to the surface.



Figure S2. Adsorption energies for the nine stable local adsorption geometries for *TCNE/Au(111)* as shown in Figure S1 applying different dispersion schemes, i.e. vdW^{surf} energies (solid bars) and MBD energies (dashed bars). As explained in the main text, we categorize the local adsorption geometries into three groups: flat lying local adsorption geometries (yellow), upright standing with the central C=C bond oriented parallel to the surface (dark blue) and equivalently normal to the surface (light blue). Qualitative different results are observed when changing from vdW^{surf} to MBD: the energy ordering for the flat lying local adsorption geometries changes, i.e. the most stable structure changes from structure G3 to G1 (but is still flat lying). More importantly for this study, the upright standing local adsorption geometries are favored when applying MBD, reducing the energy difference between structure G4 and the most stable structure G1 from 0.24 eV (vdW^{surf}) to 0.06 eV (MBD). As a consequence, the PESs in Figures S8, S10 and S12 also show qualitative different results when applying different dispersion correction schemes.



Figure S3. Number of configurations generated in Step 2 of Figure 1 obtained for different minimum distances (threshold, d_{min}) between the atoms of two neighboring molecules. The threshold chosen for TCNE/Au(111) is 2.4 Å.



Figure S4. Flowchart for the Discretization of the Configuration Space. As described in the main text the discretization is performed in two consecutive steps where we first find the local adsorption geometries isolated molecules could assume on the surface and afterwards combine them into supramolecular configurations. As illustrated in the flowchart, after evaluating all local adsorption geometries and their symmetry equivalents, all physically sensible configurations with multiple molecules per supercell are generated. In contrast to a brute-force ansatz where each molecule is put on every possible position in each possible rotation, we start with an empty supercell and consecutively add additional molecules to the supercell. To avoid collisions between the molecules we define a threshold, i.e. a minimal distance between the local adsorption geometries. On basis of this threshold, d_{min}, we either move the molecule to the next position or keep this configuration. As shown in Figure S3 the number of configurations is not overly sensitive to the exact choice of this parameter. The consecutive adding of new molecules is stopped when a certain number of molecules in the grey box) is iteratively repeated with all combinations of molecules. Thereafter, all symmetry

equivalent configurations are removed. In particular, translations as well as rotational and inversion symmetry are exploited to reduce the total number of configurations. This process of eliminating colliding structures and exploiting substrate symmetries can reduce the configuration space by orders of magnitude compared to a brute-force ansatz.



Figure S5. The 5 unique unit cell shapes obtained for TCNE/Au(111) with a coverage of three molecules and a surface area of ~ 233 Å² (31 surface Au atoms).



Figure S6. Graphical visualization for the definition of triangular structures of *TCNE/Au(111)*. For the purpose of our paper, we discarded structure that did not form equilateral triangles, i.e. where not all of these conditions were fulfilled:

- *(i) The distances between the molecules had to be equidistant (more precisely, the surface unit cells in which the molecules were placed were equidistant)*
- (ii) The molecules had to be in different symmetry-equivalent positions (i.e., rotated by 60° due to the C₃-symmetry of the molecules on the substrate).
- *(iii) The angle between the line connecting the location of the molecule within the unit cell had to be 60°.*



Figure S7. Local adsorption energies for TCNE/Au(111) calculated with different functionals. The energies are given relative to the global minimum obtained with MBD, i.e. the local adsorption geometry G3. The PBE functional in combination with TS^{surf} and MBD is compared to the SCAN functional. The SCAN functional does not lead to a re-ordering of the local adsorption geometries and shows the same trends as obtained with MBD. Remarkably, the SCAN functional even more favors the local adsorption geometry G4 which is the upright standing geometry that we find as the global minimum structure (see Figure 5 in the main text).



Figure S8 Sub-PES for TCNE/Au(111) consisting of all possible triangular polymorphs applying different dispersion schemes, i.e. vdW^{surf} (left) and MBD (right). Energies are obtained after geometry optimization and are given with respect to the global minimum of the respective PES. We see qualitative differences in the energetic ordering of the polymorphs when applying different dispersion schemes, i.e. the global minimum (indicated by the red box) changes from a structures consisting of flat lying molecules (left) to the one comprising upright standing molecules. The geometric structure of some of the polymorphs are shown in Figure S7. The boxes are arranged in the following way: starting from the center, each of the nine rings contains polymorphs consisting of three symmetry equivalent local adsorption geometries which are ordered according to their local adsorption energies as defined in Figure S2, i.e. the innermost three rings contain polymorphs comprising only flat lying molecules, while the outermost six rings comprise upright standing molecules. The ordering within the rings does not hold further meaning.



Figure S9. On the sub-PES for TCNE/Au(111) consisting of three-fold symmetric polymorphs, exemplarily, 5 different polymorphs with increasing energy are indicated and their structures are shown after geometry optimization.



Figure S10. Sub-PES for TCNE/Au(111) consisting of all possible three-fold symmetric polymorphs including one gold adatom in the center, applying different dispersion schemes, i.e. vdW^{surf} (left) and MBD (right). Energies are obtained after geometry optimization and are given with respect to the global minimum of each PES. We see qualitative differences in the energetic ordering of the polymorphs when applying different dispersion schemes, i.e. the global minimum (as indicated by the red box) changes from a structures consisting of flat lying molecules (left) to the one comprising upright standing molecules. The geometric structure of some of the polymorphs are shown in Figure S10.



Figure S11. On the sub-PES for TCNE/Au(111) consisting of three-fold symmetric polymorphs with a gold adatom in the center, exemplarily, 5 different polymorphs with increasing energy are indicated and their structures are shown after geometry optimization. The red dots indicate the positions of the gold adatom.



Figure S12. Sub-PES for TCNE/Au(111) consisting of all possible three-fold symmetric polymorphs including a vacancy in the center, applying different dispersion schemes, i.e. vdW^{surf} (left) and MBD (right). Energies are obtained after geometry optimization and are given with respect to the global minimum of each PES. We see qualitative differences in the energetic ordering of the polymorphs when applying different dispersion schemes, i.e. the global minimum (as indicated by the red box) changes from a structures consisting of flat lying molecules (left) to the one comprising upright standing molecules.