

Intelligent synthesis of magnetic nanographenes via chemist-intuited atomic robotic probe

In the format provided by the
authors and unedited

Table of Contents:

Supplementary Notes

Section 1: DFT calculations

Section 2: Additional STM/AFM/STS measurements

Section 3: Sequential cyclodehydrogenation triggered by atomic manipulation

Section 4: Determining the initial site of the cyclodehydrogenation process

Section 5: Details of Deep Neural Modules

Supplementary Figs. S1 to S36

Supplementary Tables S1 to S2

Supplementary References

Supplementary Notes

Section 1: DFT calculation

All the first-principles calculations were conducted with spin-polarized density functional theory (DFT) formalism by utilizing the Vienna ab-initio simulation package (VASP)^{63,64}. For the studies of possible reaction mechanism, including geometry relaxations and total energy calculations, the generalized gradient approximation (GGA) in the Perdew-Burke-Ernzerh (PBE) format⁶⁵, the projector-augmented wave (PAW) pseudopotentials⁶⁶ and a plane wave basis with the cut-off energy of 450 eV are employed. Van der Waals force (through DFT+D3)⁶⁷ is also included in the surface reaction simulations. The convergence criteria for electronic steps and structural relaxations were set to 10⁻⁴ eV and 0.01 eV/Å, respectively.

For the simulations of surface reaction, a 9×8 supercell of Au(111) surface with three atomic layers was adopted to model the reaction substrate, in which the bottom two layers were kept fixed in all calculations. A vacuum layer of 12 Å in the direction perpendicular to Au(111) surface was included to avoid artificial interaction between the neighbouring images. Gamma-point sampling was applied in all calculations.

In the studies of molecular orbitals (MO) for all interested surface species, the structures were kept as the relaxed geometries gained from our surface reaction simulations with Au(111) substrate removed, including initial reactant, intermediate and final product. Such a strategy is chosen by considering the following aspects: (1) The 9×8 Au(111) supercell is enormous for DFT calculations using our current computational power; (2) The distance between the molecule and Au(111) surface ranges to be ~3- 5 Å, indicating that the interaction between the molecule and surface is relatively weak⁶⁸⁻⁷⁰. In order to predict more accurate MO energy diagram, a hybrid functional, B3LYP⁷¹, was adopted in this section.

The Visualization for Electronic and Structural Analysis (VESTA) software⁷² was used for visualization of the atomic structures and molecular orbital information. The post-processing code-VASPKIT⁷³ has been utilized to produce wavefunction of the interested MOs based on DFT calculations data obtained by VASP.

Section 2: Additional STM/AFM/STS measurements

2.1 Distinguishing TA-H and molecule 1

Upon being deposited onto the Au(111) surface held at room temperature, the TA-H molecules appear to have two configurations at low surface coverages: A single molecule **TA-H** presents a non-planar appearance due to the two *sp*³ hydrogen atoms at its [3]triangulene core, as shown

in Fig. S10c and Fig. S11b . Besides, a propeller-like self-assemble structure of the as-deposited **TA-H** can be seen on the Au(111) surface, shown in Fig. S10. However, mild thermal annealing easily changes such a self-assembly pattern (thermocouple reading ~350 K), resulting in isolated molecular configurations shown in Fig. S10c and Fig. S11b. Shown in Fig. S10e and f, both an isolated **TA-H** molecule or assembled ones exhibit an edge length of ~1.5-1.6 nm. Note that the adsorption of the isolated ones on the Au(111) surface is not stable, as they can be easily moved laterally by the tip during scanning. As a result, it is challenging to perform AFM imaging over an isolated **TA-H** molecule because a reduced tip-sample distance is required to reach the Pauli-repulsion regime for a clear view of the molecule.

After mild thermal annealing at ~400 K (thermocouple reading, maximum is 425 K), the two pristine absorption configurations mentioned above vanish, leading to the formation of the isolated products on the Au(111) surface, which is named molecule **1** shown in Fig. 2a in the main text and Fig. S10d. Compared to the molecules prior to the mild annealing, the molecules after annealing (molecule **1**) exhibit a relatively stable adsorption configuration on the Au(111) surface under the scanning with different bias voltages. The edge length of an isolated molecule **1** is ~1.6 to 1.7 nm, close to that of **TA-H** before annealing (Fig. S10c and Fig. S11b).

In order to exclude different adsorption configurations that may affect the STM topography, lateral manipulation of molecule **1** on the Au(111) surface is performed (Fig. S12). During the movements, the feature of molecule **1** captured by STM imaging did not change, showing the same feature as the one shown in the STM images right after the mild annealing process. In addition, the STM topography of molecule **1** on the elbow site of herringbone reconstruction appears to be the same as on the *fcc* site, showing that the adsorption configurations do not affect the STM topography.

Furthermore, we performed additional DFT calculations to relax the TA-H molecule on the Au(111) surface. Considering that the *sp³* hydrogen atoms might have different configurations (e.g., protruding upward away from the surface, protruding downward towards the surface, and alternative configuration), we calculated all possible adsorption configurations of TA-H molecule on Au(111) surface. Our DFT calculation results reveal that the TA-H molecule appears to have minimum total energy when the *sp³* hydrogen atoms protrude downwards towards the Au(111) surface. With the fully-relaxed TA-H molecule on the Au(111) surface with minimum total energy configuration, we conducted simulated AFM imaging using the probe particle model (PPM) and compared the results with the simulated one of molecule **1**. As shown in Fig. S8, the significant differences between TA-H and molecule **1** in relaxed molecular structures and simulated AFM images can be revealed. Due to the two *sp³* hydrogen

atoms protruding downward to the surface, one of the corners of the molecule is levelled upward, resulting in a strong contrast at one corner and an asymmetric feature over the molecule in the simulated AFM image (Fig. S8f), differing from our experimental AFM image (Fig. S8c) and the simulated AFM image of substituted [3]triangulene (molecule **1**, Fig. S8b). Such a result can be further supported by the STM measurements using a CO-functionalized tip, which is sensitive to the variations in the z-direction. As shown in Fig. S11, the STM image of an isolated molecule **1** on Au(111) surface captured by a CO-tip exhibits a relatively homogenous and centrosymmetric feature rather than the existence of a tiled feature, indicating that there are no protruding *sp*³ hydrogen remaining on the [3]triangulene core (If *sp*³ H exists, an asymmetric with tilted feature can be captured by CO-tip).

The minimum energy needed to break the *sp*³ C-H bond in a **TA-H** molecule is calculated, showing an energy value of ~0.9 eV. Such a value shows a good agreement with the experimental temperature window (375 K-425 K).

2.2 Observing Kondo resonance feature of molecule **1** and molecule **3** on Au(111) surface

The small-range dI/dV measurements show a zero-bias peak over molecule **3** on Au(111) surface (Extended Data Fig. 3), presumably arising from partly screened spin an *S*=1 Kondo resonance. However, such a feature is absent over a molecule known as substituted [3]triangulene with *S*=1 on Au(111) surface (Fig. S17).

The Kondo effect might manifest as a resonance peak or dip centred at the Fermi level, which could be probed by performing STS measurement below the Kondo temperature ($T < T_K$). The Kondo temperature is related to several factors, including magnetic impurity energy level (Γ) with respect to Fermi level (E_F), on-site Coulomb repulsion energy (U), and the hybridization between impurity state and host⁷⁴⁻⁷⁶. A recent report on a non-substituted [3]triangulene on Au(111) surface exhibits the observation of a weak Kondo feature at 4.5 with an estimated Kondo temperature range of $4.5K < T_K < 14$ K, which is due to the underscreened spin on the metal surface⁷⁷.

By comparing molecule **1** with the non-substituted [3]triangulene reported recently, the absence of the Kondo feature could be ascribed to the three *m*-Xylene side groups of molecule **1** that lift the centred [3]triangulene core, highly reducing the hybridization and screening between the [3]triangulene core of molecule **1** and Au(111) surface, which could further reduce the Kondo temperature of the substituted-[3]triangulene system⁷⁸.

Shown in Fig. S9, it can be observed that the distance between molecule **1** and Au(111) surface is 5.07 Å, much larger than that of non-substituted [3]triangulene molecule (3.15 Å). Note that 5.07 Å can be regarded as a weak-physical-adsorption distance that might significantly reduce the hybridization and screening between the molecule **1** and Au(111) surface underneath, possibly resulting in difficulty in observing the Kondo resonance features at 4.5 K because a Kondo temperature range of $4.5\text{K} < T_K < 14\text{ K}$ for a non-substituted [3]triangule is estimated in a recent report. Besides, such a claim could be further supported by the observation of the Kondo feature over molecule **3** presented in Extended Data Fig.3. As shown in Fig. S7, the calculated distance between molecule **3** and Au (111) surface ranges from 3.17-4.38 Å, similar to that of non-substituted [3]triangulene. Such a reduced molecule-substrate distance results from the planarization promoted by probe-assisted cyclization.

Section 3: Sequential cyclodehydrogenation triggered by atomic manipulation

Supplementary Fig. S18-19 show additional information on the stepwise cyclization of molecule **1-3** on Au(111) surface. We identified the proper region to position the STM tip for a high chance of the successful cyclodehydrogenation of a molecule **1**, **2**, and **3** on Au(111) surface. Specifically, we tested different regions over/nearby the molecules, the reaction can be triggered when positioning the tip over the region around the zigzag edge and protruding CH₃ groups, as shown in Fig. S18a.

Prior to the manipulation, the tip was parked over the identified region with a setpoint of $V_s = 0.5\text{ V}$, $I_t = 400\text{ pA}$. Next, the feedback loop was switched off for the operation in constant-height mode. Then, the bias voltage V_s was set to slightly above 2.70 V and kept for 5-10 seconds. Typically, the tunneling current increases up to ~20 nA and stayed at this value for a few seconds before it falls sharply to the original setpoint (Fig. S19b), indicating the accomplishment of a manipulation. An STM or an AFM image will be taken to check the manipulation outcome. It is noted that the voltage pulse may cause a change of the adsorption geometry or lead to the formation of undesired non-cyclized products, cyclized with -CH₂ termini, and even directly picking up the molecules under certain circumstances (Fig. S20).

To investigate the threshold of the voltage pulse, we performed bias-dependent experiments at different voltages (V_s). For each manipulation, we kept the applied V_s up to 180 seconds at constant-height mode, and simultaneously monitor the tunnelling current (I_t). As mentioned above, a sharp current drop usually indicates a successful manipulation, and the product can be confirmed by STM imaging. If a current drop does not occur within 180 seconds, the applied

voltage V_s will be regarded as below the threshold. In addition, to further confirm the reliability of the results, different tip-sample distances (i.e., different setpoint I_t prior to opening the feedback loop, and the initial bias voltage V_s is always set as 0.5 V) were examined.

Since constant-height mode was applied during the manipulation, the tunneling current I_t varies slightly over the time. Thus, we averaged the tunneling current I_t over the manipulation time t . In this way, the number of electrons for a manipulation case can be estimated by multiplying the averaged tunneling current I_t and the manipulation time t . Thus, the average yield η can be estimated as $\frac{e}{I_t \times t}$ (e refers to the elementary charge)⁷⁹. As shown in Fig. S19a, the lowest bias voltage to trigger the cyclodehydrogenation process is determined to be 2.40 V. Below 2.40 V (e.g., 2.30 V), the reaction cannot be triggered even at a very closed tip-sample distance (e.g., $I_t = 1.6$ nA or 2.0 nA before switching off the feedback loop).

The current trace recorded during the cyclodehydrogenation process reveals that I_t usually undergoes a sudden increase upon applying the bias voltage (usually above 2.70 V). A sharp current drop will occur after a few seconds. This step-like current trace is typically observed during the first cyclization process (i.e., molecule **1** to molecule **2**, Fig. S19b). In contrast, an additional spike feature can be frequently observed before the current drops during the second (molecule **2** to molecule **3**) and the third (molecule **3** to **4**) cyclization process (Fig. S19c-d, as highlighted by green arrows).

Intuitively, such an observation may be interpreted as the distance reduction of the tunneling junction, which could be induced by the slightly vertical displacement of either the target molecule or the tip. Note that the manipulation process is conducted under the constant-height mode, which indicates that the vertical position of the tip (z-direction) will not be affected. Therefore, it is reasonable to assume the tip holds its z position after switching off the feedback loop. In this regard, the spike feature is likely attributed to the vertical motion of the target molecule in a short period of time during the cyclization process, consistent with our DFT-calculated adsorption configuration of reaction intermediates (Fig. 5 in the main text). The *m*-Xylene moiety will rotate by ~30 degrees towards the C-C coupling, altering the distance between the target molecule and the tip, presumably leading to a spike feature in the tunneling current trace. The coupling of carbon atoms to the adjacent sigma carbon radicals on the zigzag edge creates the planar molecules on Au(111), resulting in a sudden drop in the tunneling channel.

It is worth noting that the spike feature is less frequently observed during the cyclization of molecule **1**. After the first cyclodehydrogenation process, the molecules are significantly bent due to different heights on Au(111) surface compared with molecule **1**(Fig. S7). The bending could result in a substantial intramolecular strain in molecule **2** and **3** that may hinder the rotation of *m*-Xylene moieties during the cyclization process. Thus, when conducting a second and a third cyclization process, molecules might release intramolecular strain during the rotation of the *m*-Xylene moieties, resulting in the spike features in the tunneling current trace. In contrast, molecule **1** is a three-fold symmetry molecule with larger distance to the Au(111) surface (Fig. S7), and its three *m*-Xylene moieties are supported nearly equivalently on Au(111) surface. Thus, molecule **1** is more flexible for the rotation of the *m*-Xylene moiety towards the cyclization. Thus, the spike feature is less frequently observed in the current trace during the process from molecule **1** to molecule **2**.

Section 4: Determining the initial site of the cyclodehydrogenation process

As discussed in the main text, the cyclodehydrogenation process triggered by the tip can be subdivided into a few steps. Determining the initial reaction site also helps to understand the reaction mechanism. Direct observation from the STM image remains challenging to identify the initial site, as the phenyl -CH site from core [3]triangulene and the benzylic -CH site from the *m*-Xylene head are closed to each other, resulting in the appearance of a blurred boundary in the STM image (see Fig. S18a). In this regard, we performed DFT calculations for different circumstances to figure out the initial reaction site of this process (Fig. S6).

We first consider the phenyl -CH as the initial site. From the results shown in Fig. S6a, by removing two hydrogen atoms from the phenyl core, the molecule adopts a stabilized adsorption on the Au(111) surface. A similar situation also can be observed from the calculation results of removing two hydrogen atoms from each benzylic -CH site that can lead to a stable adsorption on the Au(111) surface (Fig. S6b). Such results imply that cyclodehydrogenation cannot be promoted by simply breaking the C-H bonds from phenyl or benzylic -CH sites.

Next, we examine other possible factors that may affect the process. It can be observed that there is an apparent steric hindrance existing between the phenyl -CH and the adjacent -CH₃

caused by the repulsion of hydrogen atoms, leading to a tilted configuration of all *m*-Xylene moieties of molecule **1** (Fig. S7a and S7d). Hence, the removal of hydrogen atoms would also relax the steric hindrance by allowing the *m*-Xylene moiety to rotate and thus to adopt the new energetically favorable adsorption configuration on the Au(111) surface. In this regard, the rotation of the *m*-Xylene moiety needs to be considered in the DFT calculations.

Fig. 5 in the main text shows that the reaction can be promoted after removing phenyl *sp*2 hydrogen atoms and rotating the *m*-Xylene moiety for ~30 degrees, yielding a cyclized product. Therefore, the cyclodehydrogenation is triggered from the phenyl -CH site of the [3]triangulene core. Similarly, Extended Fig. 4a and 4b present the second (molecule **2** to **3**) and the third (molecule **3** to **4**) cyclization process with corresponding energy diagrams.

Section 5: Details of Deep Neural Modules

5.1 Model architecture for faster R-CNN (Target Identification, TI module)

The prediction model aims to realize the position detection and classification of the target molecule based on the Faster Region-based Convolutional Neural Network (Faster R-CNN), which is a deep learning-based object detection architecture widely used in the machine learning (ML) field for target object detection and classification. The general block diagram of the architecture of the Faster R-CNN model is shown in Fig. S21. The faster RCNN model consists mainly three important modules named fast region-based convolutional neural networks (FPN), regional proposal network (RPN), and region of interest (ROI) head.

The FPN extracts feature maps of the different scales from input STM images and, hence, enables Faster R-CNN to detect molecules of various sizes, revealed by STM images with different resolutions. The FPN backbone is constructed by a total of fifty convolutional and ReLu layers (ResNet-50)⁸⁰. The detailed architecture of ResNet-50 is shown in Fig. S31. In our automated system, a large-scale STM image is in a shape of [3, H, W], where three channels are used to store colour information (red, green, and blue), H is the height of the image, and W is the width of the image. This image will be input into the ResNet-50, and features with different scale will be extracted from different positions of the ResNet-50. In our case, five different feature maps are extracted and as the level of feature maps increases, the scale and resolution of feature maps decrease. In the meantime, the spatial coverage and abstract information of feature maps gradually increase. The downstream modules then can use a mixture of different level of features to realize the target molecule classification and

localization.

The RPN module generates bounding boxes (a rectangular box containing location information), or regional proposals, on the extracted feature maps from FPN and gives a score for each bounding box to contain objects of interest. After this stage, bad quality bounding boxes (i.e., low score, out of the region, etc.) are removed. In addition, duplicate anchors are also removed by the non-maximum suppression (NMS) method which is introduced in the latter text. The proposal information from RPN module is combined with the raw feature information from FPN and passed into the ROI module.

The ROI module will convert the information from FPN and ROI into fixed-sized vectors, and then the fixed-sized vectors are passed into the fully connected layers to output two different matrixes: 1) A three-channel matrix containing the scores for the proposals to be each of the three classes (target molecules, non-target molecules, or a background); 2) A four-channel matrix containing four final bounding box fine-tuning parameters. The probability distribution for a detected molecule to be the target molecule is obtained from the scores by using the SoftMax function.

5.2 Model architecture for key point faster R-CNN (Site Justification, SJ module)

The key point faster R-CNN model is a modified form of the faster R-CNN model. Apart from all the modules of faster R-CNN model, there is one more module called the key point head which can output the key point mask which is the exact location of the reaction sites in our case.

5.3 Model architecture for product classification CNN (Product Classification, PC module)

The product classification module is a straightforward combination of convolutional neural networks and fully connected neural networks. The image is firstly processed into various features by the convolutional layers and then these features are passed into the fully connected layers to do the classification task. The probability distribution for the product to be ‘no reaction’, ‘successful reaction’ or ‘failed reaction’ is obtained from the scores by using the SoftMax function.

5.4 Information exchange and interface platform

To bridge the deep neural modules and the controlling system of the scanning tunneling microscope, a LabVIEW-based interface is developed. It can conduct information exchange and real-time operations according to the decision made by the deep neural modules. In this

STM-AI system, LabVIEW Python Node is utilized to connect the three deep neural modules and LabVIEW interface, and LabVIEW Nanonis VI, a package provided by SPECS GmbH, is applied to link the controlling unit of the scanning tunneling microscope and the interface (Fig. S24-S30).

Fig. S24 illustrates the overall construction design of the STM-AI system. A $30\text{ nm} \times 30\text{ nm}$ frame will be scanned over the surface when the manipulation program starts (Parameters: $V_s = 0.5\text{ V}$, $I_t = 500\text{ pA}$). The as-scanned STM image is captured and passed to the TI module. Meanwhile, the TI module is applied to detect the target molecule **1** and record their locations. If no molecule **1** is detected, the LabVIEW interface will automatically move to the next frame until a target is detected. After detecting molecule **1**, a magnified $4\text{ nm} \times 4\text{ nm}$ frame will be scanned over the recorded target location list to get a clear and stabilized STM image for a single molecule **1**. Then the single-molecule STM image will be transferred to the SJ module to justify the possible reaction site, and a voltage bias will be applied to the justified site for 40 seconds with an open-feedback loop as the manipulation process. Upon the manipulation accomplishes, the feedback loop will be closed, and the scanning parameters will reset as the values before manipulation automatically. Followed by a quick scan, the STM image will be passed to the PC module to judge whether the manipulation is successfully applied. Note that if ‘no reaction’ is classified, the manipulation will repeat up to three times until a reaction is triggered (Sometimes the drift might not be fully minimized, leading to a ‘no reaction’ outcome). If a reaction is triggered, the PC module will judge whether the cyclization is appropriately promoted by the manipulation and take the record. If a reaction is triggered with either failed or successful outcomes, the system will move to the next molecule in the target list or continue target searching in large-scale. After all the molecules in the target list are processed in sequence, the TI module will take over the STM and repeat the target searching/identification step in the next $30\text{ nm} \times 30\text{ nm}$ frame.

5.5 Definition of average precision (AP) and average recall (AR)

When evaluating the performance of the inferencing results of the TI model, we have used the average precision (AP) and average recall (AR) metrics. To get values of precision and recall, an intersection over union (IoU) threshold is required. The equations for precision and recall are defined as following:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{S(1)}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{S(2)}$$

where TP stands for a true positive, FP stands for a false positive, and FN stands for a false negative.

A true positive requires a correct classification result and an IoU value between the real position of the target molecule in the STM image and the detected position higher than the threshold (i.e., an IoU thresh-old of 0.8 means a bounding box 20% larger or smaller than the real target molecule on STM image will be still counted as a false positive). Therefore, recall-IoU and precision-recall curves can be drawn by setting the threshold for IoU from 0.5 to 1.0. And the average recall and average precision can be computed as the average over all $\text{IoU} \in [0.5, 1.0]$. It can be calculated as two times the area under the Recall-IoU curve as shown in Fig. S32a.

$$\text{AR} = 2 \int_{0.5}^1 \text{Recall}(\text{IoU}) d(\text{IoU}) \quad \text{S(3)}$$

For the Precision-Recall curve, the interpolated precision at a certain recall level is defined as the highest precision found for any recall level. AP can then be defined as the area under the interpolated precision-recall curve as shown in Fig. S32b:

$$\text{Precision}_{\text{interpreted}}(\text{recall}) = \max_{r' \geq \text{recall}} \text{Precision}(r') \quad \text{S(4)}$$

$$\text{AP} = \int_0^1 \text{Precision}_{\text{interpreted}}(\text{recall}) d(\text{recall}) \quad \text{S(5)}$$

5.6 Recognition Rate

The recognition rate is defined as the percentage of correctly recognized molecules from visual inspection. Since the number of molecules in training data and test data is limited in this demonstration case, the chemists were able to correctly label the ground truth by spending a long time and double-checking the results, so that an exact recognition rate can be obtained for comparison.

5.7 Reaction Site Recognition Rate

For the SJ module, the reaction site recognition rate (RSRR) is used as the performance metric. It utilized the similar idea that transforms an ambiguous continuous problem into a clear discrete problem as the Percentage of Detected Joints metric used for human pose estimation. As shown in Fig. S33, the detected reaction site is considered correct if the distance between the predicted and the reference sites is within a certain fraction of the bounding box diagonal of the target molecule. We use an annular region with a diameter of 2 Å (~12.5% of the edge length of the molecule), centred at the reference sites proposed by expert knowledge. If the

predicted sites locate within the region, thus it would be judged as correct site.

5.8 Non-maximum suppression

It works as the bounding box with the highest score is retained when multiple bounding boxes have a large intersection over union (IoU). The above circumstance indicates those bounding boxes probably cover the same molecule. RPN then outputs the remaining bounding boxes with high scores as proposals. Each proposal is a bounding box obtained by preliminarily fine-tuning, it likely contains a molecule. However, most bounding boxes would be discarded as duplicates or backgrounds at this stage.

5.9 Intersection over union (IoU) for NMS operation

IoU measures the overlapping between two bounding boxes. As shown in Fig. S34, a large IoU value indicates two bounding boxes contain the same molecule. Therefore, only one bounding box is required, and the bounding box with a lower confidence level should be discarded.

5.10 Mechanism of SHapley Additive exPlanations (SHAP) method

The SHAP analysis is a game theoretic approach to explain the output of any machine learning model (including the complex deep learning models). It can find the feature importance for each input inside the target ML models, thus enabling the users to interpret the models. This feature importance finding process is similar to the process in finding the contribution of each player in a collaborative game. A simplified example is presented below to illustrate how the feature importance is calculated in SHAP analysis.

As shown in Extended Fig. 5, there are three players A, B, and C. They collaborate with each other to play a game. When all of them have joined the game, based on their different skills points in a specific game (e.g., 6, 4 and 4 for players A, B, and C.), they can achieve a 100 reward. The task is to quantify how important each player is in getting the reward.

To calculate the contribution of each player, a specific sequence (e.g., player A first, then player B, next player C) of how each group member joined the game need to be assumed and the marginal reward of each player is recorded. In this sequence, player A is the first member to gain a reward of 45, then player B joins the game and brings the reward to 85, and next player C joins to bring the reward to 100. Therefore, the players' respective marginal rewards would be (player A, B, C = 45, 40, 15).

However, the calculated marginal reward for a specific sequence may not accurately represent the contribution of each player. When the sequence is changed from player A, B, C to player

A, C, B, if players B and C have a similar skill set, the rewards are still 45, 85, and 100 after player A, C, B joins the game, respectively. Then the players' respective marginal reward is changed (player A, B, C = 45, 15, 40). Therefore, the sequence of how the players join the game is important.

In order to get a more accurate reward of each individual player, the marginal reward of each player needs to be calculated under every possible sequence. The reward for each individual player then can be the sum of these marginal rewards over the number of possible sequences (the calculation for a specific player is illustrated in Equation S6). For example, in the case outlined above, we would simulate all the sequences: ABC, ACB, BCA, BAC, CAB, and CBA, and the marginal reward of each player is recorded for each sequence. Then, by averaging all these rewards, the reward contributed from each player can be obtained. This reward is the SHAP value.

SHAP value of a player =

$$\frac{\text{sum of marginal reward of this specific player under all possible sequence}}{\text{number of total possible sequence}} \quad \text{S(6)}$$

Back to the feature importance analysis of our three deep neural modules, we can take the problem as an analogue to the above case (Extended Fig. 5a). Since the number of pixels in one image is a huge number (i.e., about 495616) in our cases, it is impossible to permute all possible sequences for such a huge number of features. Instead, we have utilized feature blocks (i.e., as shown in Extended Fig. 5b, an image is segmented into fifty segments as feature blocks) for the SHAP analysis. As a result, all the feature blocks in an image are regarded as players, which are fed into the deep neural modules to obtain the predictions for each task (i.e., the bounding boxes of target molecules for TI module, the reaction site position for SJ module, and the label of the molecule for PC module.). The prediction process is treated as the game. The average precision for TI and SJ modules, and the cross-entropy loss for PC module are treated as the reward. With a similar calculation of Equation S7, the SHAP value of each feature block in an image can be calculated, and this value is used to measure the importance of the feature blocks.

SHAP value of a feature block =

$$\frac{\text{sum of marginal reward of this feature block on the prediction tasks under all possible sequence}}{\text{number of total possible sequence}} \quad \text{S(7)}$$

The above process is the interpretation of the prediction model on a specific data point which is called the local interpretation. To get a more global interpretation of the deep neural modules, we can plot the SHAP values of every feature block for a variety of representative data points as shown in Fig. S35. The open-source code to implement SHAP analyses in Python is provided

in GitHub (<https://github.com/jiali1025/Intelligent-topological-engineering-of-quantum-p-magnets>).

5.11 Examples of SHAP analysis of deep neural modules

Examining the training outcome of the deep learning modules of the STM-AI system by using SHAP analysis reveals that each module adopts different strategies to grasp the crucial information through the dataset training. As illustrated in Fig. S35, it can be revealed that the recognition of molecular precursors (TI) and the classification of reaction products (PC) majorly rely on accurately identifying the geometrical topologies of the targets, indicating that the machine-vision-based tasks drive the TI and the PC modules. Such a conclusion can be directly validated by examining the SHAP analysis results shown in Fig. S35 (a to c, g to h, right panels), as the relatively-high-value regions cover the whole areas of the molecular precursors and products, exhibiting similar outlines depicted by STM imaging. In contrast, the critical information extracted *via* the SJ module is not identical to the geometrical outlines imaged by STM, indicating that such a module is not simply driven by a machine-vision-based task but possibly a physical-chemistry-related task. Detailed analysis of the SJ module is presented in the main text and Fig. 4.

5.12 Testing the current CARP by handling unseen molecules

The current version of CARP in the manuscript is designed to perform site-selective cyclodehydrogenation along the zigzag edge directions that can synthesize nanographenes with tailored geometric and magnetic configurations using supervised-learning-based learning paradigm (e.g., The synthesis from molecule **1** to molecule **2**, as discussed in the manuscript). In principle, additional database training is required to extend the knowledge of the deep neural networks to perform different reactions over unseen molecular precursors based on the current CARP framework, which is determined by the methodology. However, it is necessary to evaluate and compare the performance of CARP when handling unseen molecules.

In this regard, additional tests with different strategies are conducted on the SJ module of CARP because the accuracy of site judgement is crucial to a successful reaction during the practical operations of the AI-operated STM, which determines whether the CARP framework could apply to other new structures and unseen systems. Regarding this, molecule **2** and molecule **3** were utilized to test the SJ module because they were reported for the first time in the manuscript, which is undoubtedly unseen to the deep neural networks.

First, the SJ module of current CARP is tested by molecule **2** and molecule **3** to justify the possible reaction sites of the cyclodehydrogenation to be conducted. As shown in Fig. S36a, the justification accuracy is 59.44% for molecule **2** and 16.17% for molecule **3**, respectively. Compared to the accuracy data presented in Fig. 3 in the manuscript, the current SJ module exhibits barely satisfied accuracy over molecule **2** while unsatisfied performance over molecule **3**. The relatively high accuracy of molecule **2** can be regarded as an exception because the molecular geometric configuration of molecule **2** is highly identical to that of molecule **1**. However, as for molecule **3**, with a different molecular geometric configuration from molecule **1**, the accuracy drops dramatically. Such discrepancies can be ascribed to the learning paradigm of supervised learning (i.e., learn from labelled train data sampled from certain distributions, when the test distribution differs a lot from the train distribution, the performance will be unsatisfied.) applied in the current version of the CARP framework. Therefore, adding additional database training to the SJ module is the best strategy.

Next, additional training sessions are performed for the SJ module with a limited volume database (50 STM images for each molecule) that contains expert knowledge on molecule **2** and molecule **3**. As shown in Fig. S36b, the site justification accuracy is significantly improved (84.91% for molecule **2** and 85.71% for molecule **3**, respectively). Note that the knowledge-enhanced CARP can be efficiently trained to achieve a high recognition accuracy and data efficiency, implying that under the current deep learning framework, the CARP can conduct operations over unseen molecules with an additional fast training process.

5.13 Original STM images and target-predicted STM images by deep neural modules

Table S2 summarizes the comparison between labelling results by the chemist and deep neural modules, including target molecule labelling, reaction site labelling, and product labelling. Comments are presented in right column of the table.

5.14 Comparison between human-operated STM and AI-operated STM (CARP)

The realization of AI-operated STM (CARP) to conduct autonomous site-selective single-molecular reactions is powered by well-trained deep neural networks. Note that the database of knowledge/experience used to train the deep neural networks originated from the human-operated STM operations. By adopting the information from human experts, the AI-operated STM's ultimate goal is to perform real-time STM operations to conduct single-molecule reactions as sophisticated as human-operated STM performance. Therefore, it is crucial to

elucidate that the nature and outcome of such reactions remain consistent, irrespective of whether human-operated STM or AI-operated STM manages it.

During the proof-of-concept test, the CARP conducted 50 reactions and 26 products (molecule **2**) were successfully synthesized, compatible with the manual operation results. The reaction success rate for manual operation is ~50%-60% due to the complexity of multi-bond-involved reactions. In this regard, the CARP operation is comparable to experienced human operators, exhibiting the potential to perform long-time operations with stable performance.

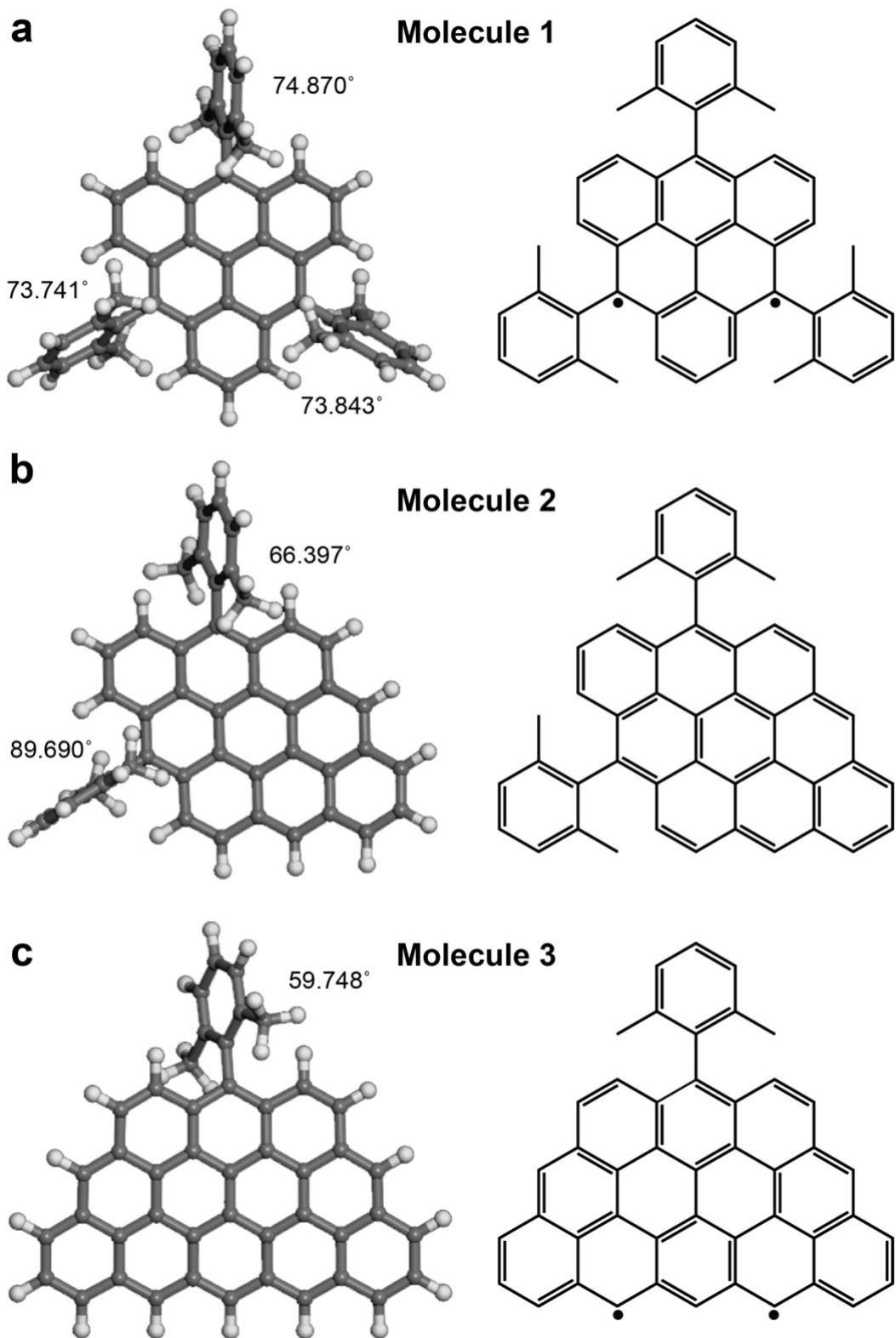


Fig. S1. DFT-optimized molecular structure of molecule 1 (a), 2 (b) and 3 (c) on Au(111) surface. Note the substrate was removed for a better clarity. The angle between each plane of *m*-Xylene moiety and the plane of the triangular core in the centre is indicated on the left panel. The grey and white balls represent carbon and hydrogen atoms, respectively.

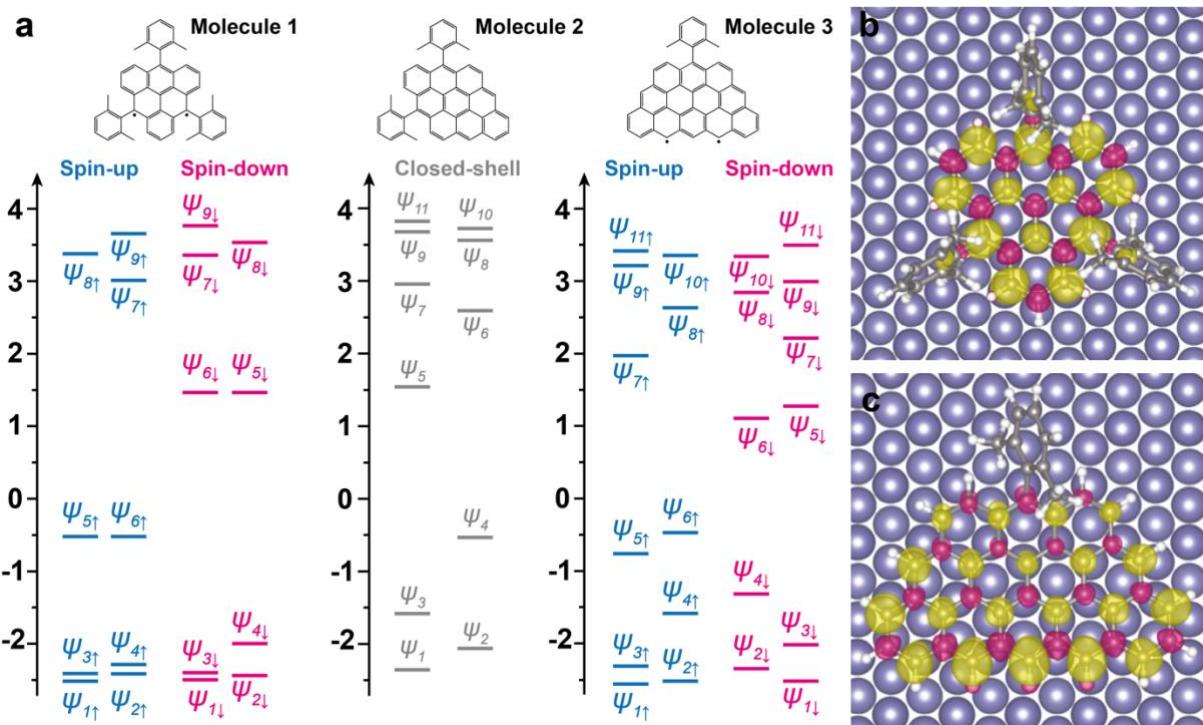


Fig. S2. DFT-calculated energy diagrams of molecule 1-3 and spin density of molecule 1 and molecule 3. (a) DFT-calculated energy diagrams of molecule 1 (open-shell triplet), molecule 2 (closed-shell), and molecule 3 (open-shell triplet). (b-c) Calculated spin density distribution of molecule 1 and molecule 3 on Au(111) surface. Yellow and pink represent spin-up and spin-down, respectively.

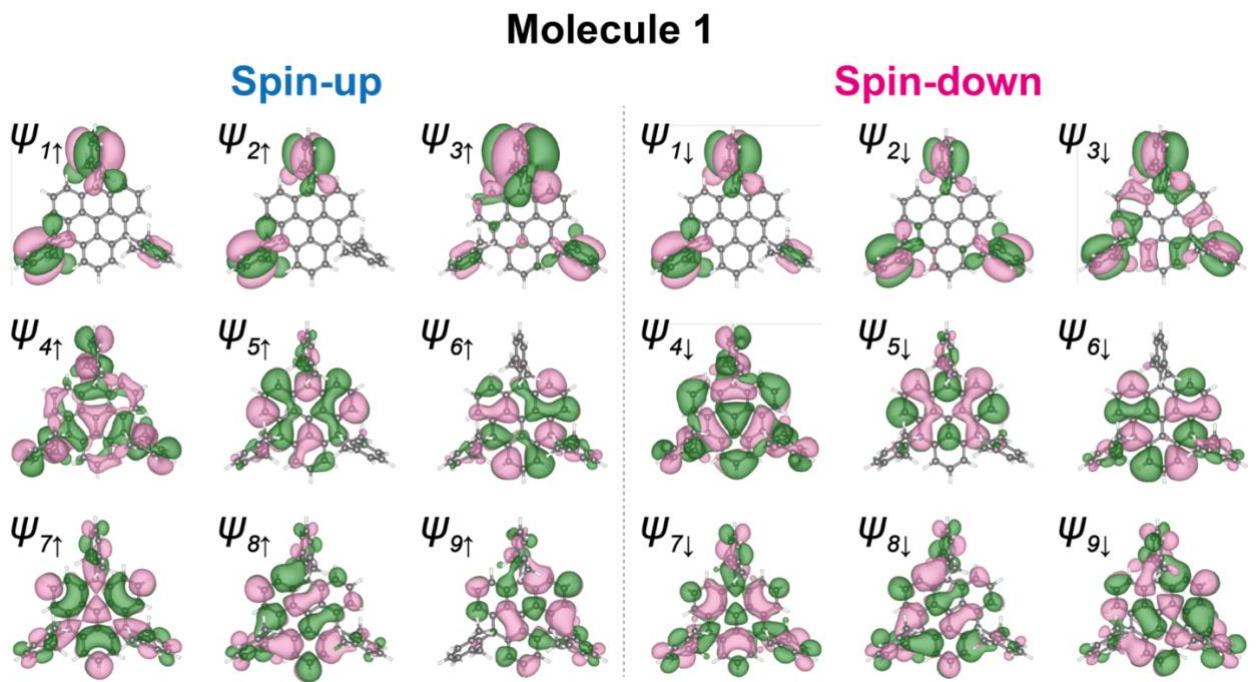


Fig. S3. DFT-calculated wave function patterns of the molecular orbitals of a freestanding molecule 1 at its open-shell triplet ground state.

Molecule 2
Closed-shell

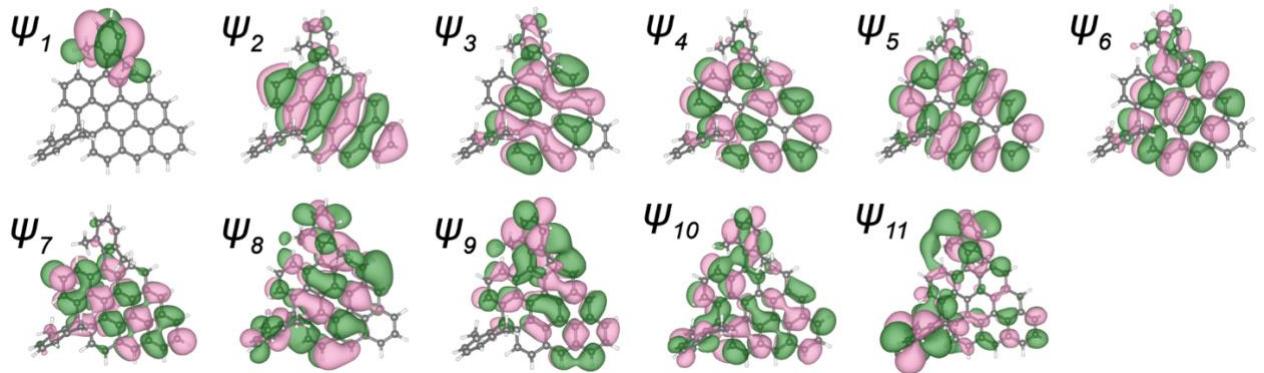


Fig. S4. DFT-calculated wave function patterns of the molecular orbitals of a freestanding molecule 2 at its closed-shell ground state.

Molecule 3

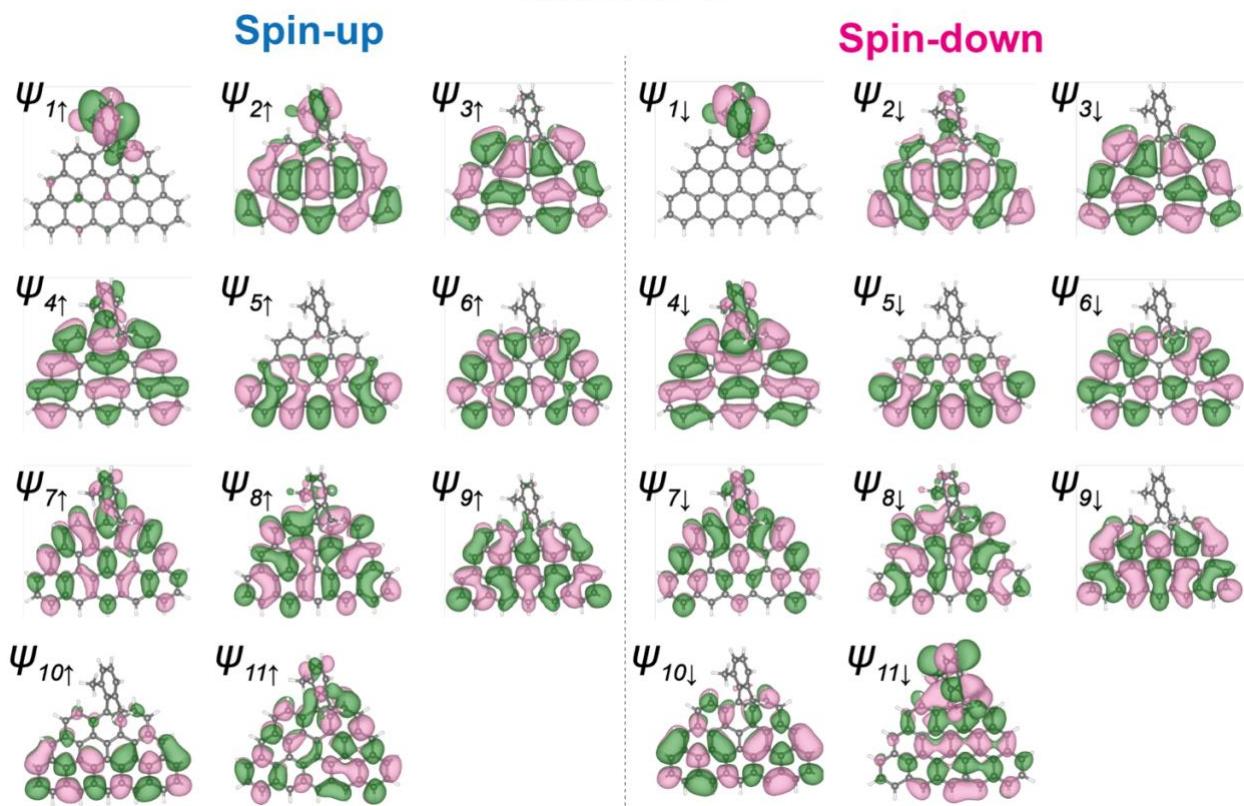


Fig. S5. DFT-calculated wave function patterns of the molecular orbitals of a freestanding molecule 3 at its open-shell triplet ground state.

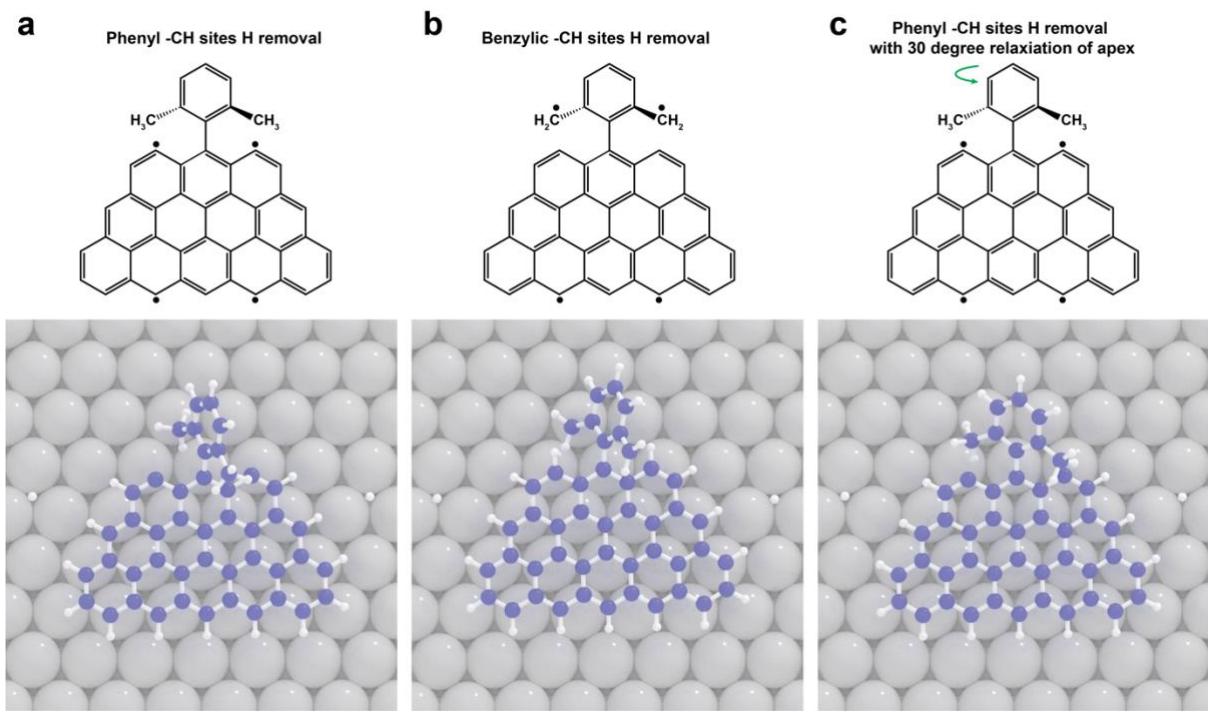


Fig. S6. Two possible pathways to initiate the cyclization are considered in the DFT calculation (Using molecule 3 to demonstrate). (a) Dehydrogenation starts from *sp*² phenyl -CH site. (b) Dehydrogenation starts from *sp*³ benzylic -CH site. (c) Dehydrogenation starts from *sp*² phenyl -CH site with a rotation of *m*-Xylene moiety for 30 degrees. The purple and white balls represent carbon and hydrogen atoms, respectively.

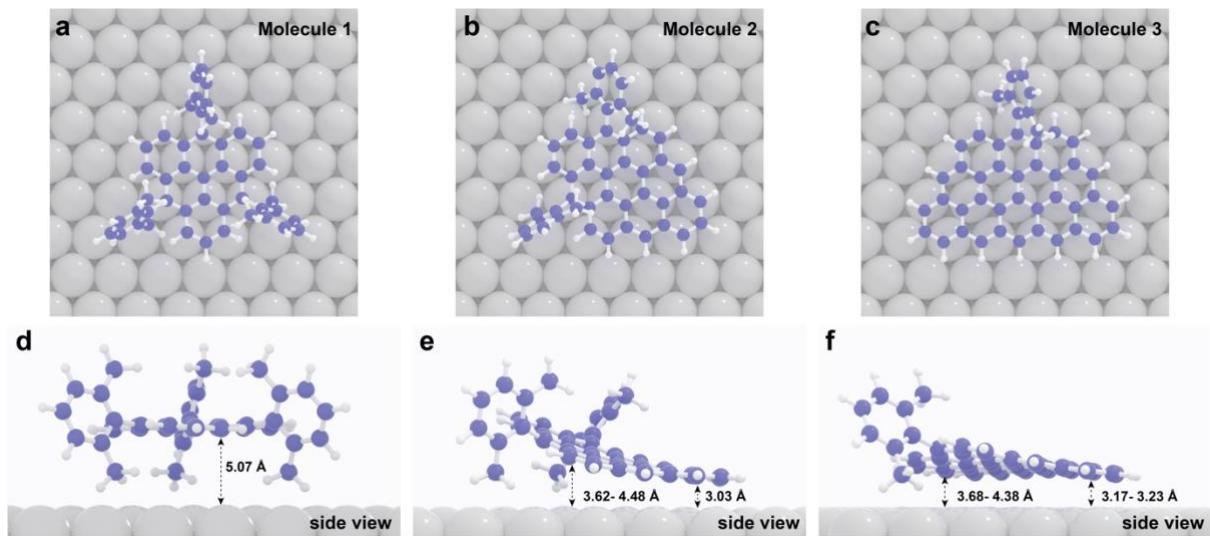


Fig. S7. DFT-calculated molecular structures relaxed on Au(111) surface. (a-c) Top-view of the structure of molecule **1**, molecule **2**, and molecule **3** on Au(111) surface, (d-f) Corresponding side-view of (a-c). The purple and white balls represent carbon and hydrogen atoms, respectively.

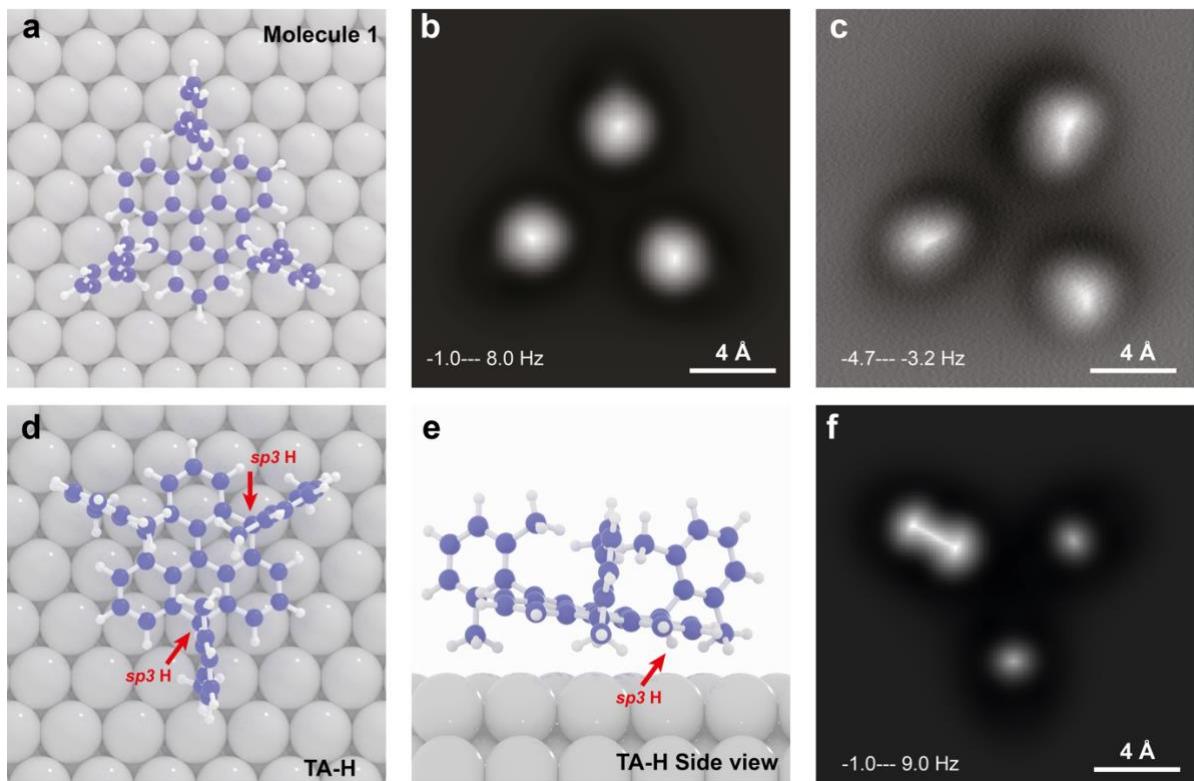


Fig. S8. Structural comparison between TA-H and molecule 1. (a-c) DFT-relaxed structure, simulated AFM image, and experimental AFM image of Molecule 1, respectively. (d) DFT-relaxed structure of TA-H. (e) Simulated AFM image of TA-H molecule. (f) Experimental STM image of molecule **1** on Au(111) surface. Scanning parameters: (c): Set point prior to turning off feedback: $V_s = 0.5$ V, $I_t = 400$ pA, scanning bias voltage: 1 mV.

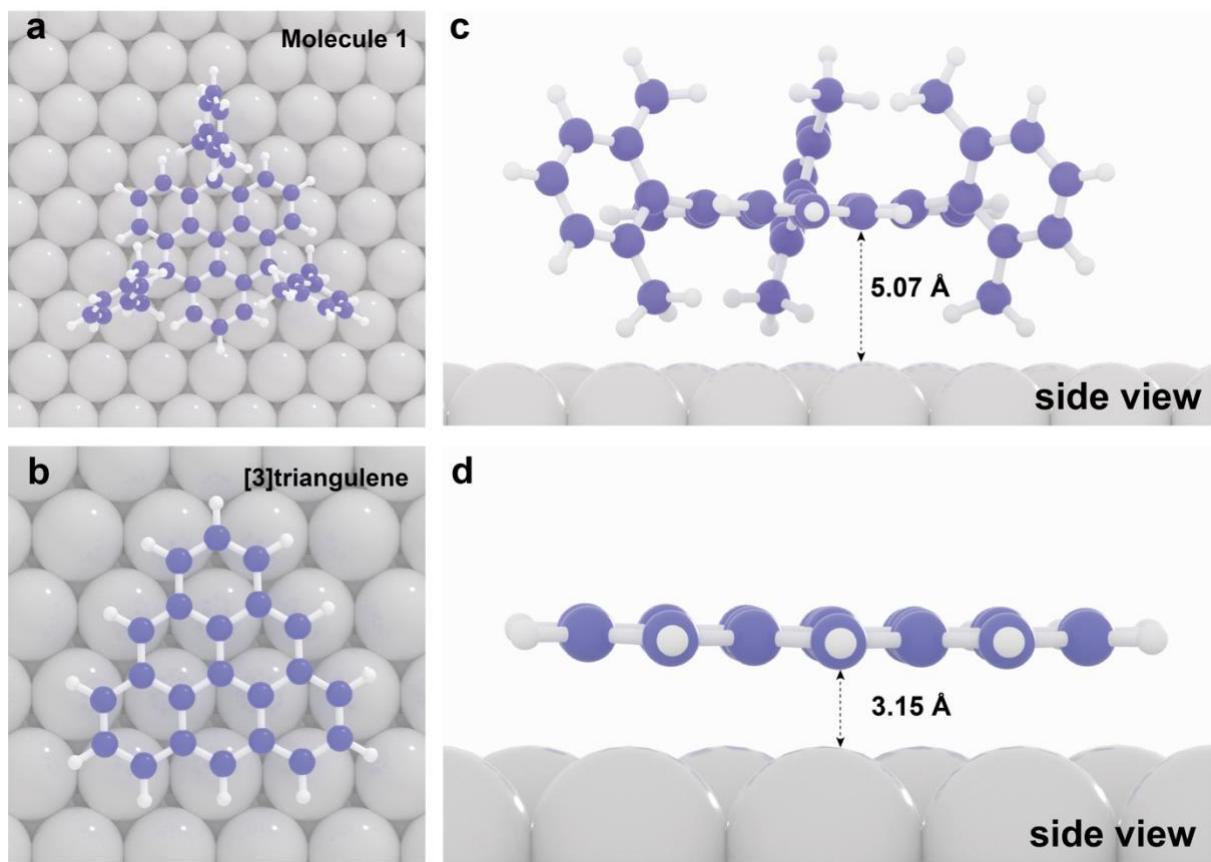


Fig. S9. DFT-relaxed structural comparison between molecule 1, non-substituted [3]triangulene. (a-b) Top view and side view of a molecule 1, and [3]triangulene relaxed on Au(111) surface, respectively. **(c-d)** Side view and side view of a molecule 1 and [3]triangulene, relaxed on Au(111) surface, respectively.

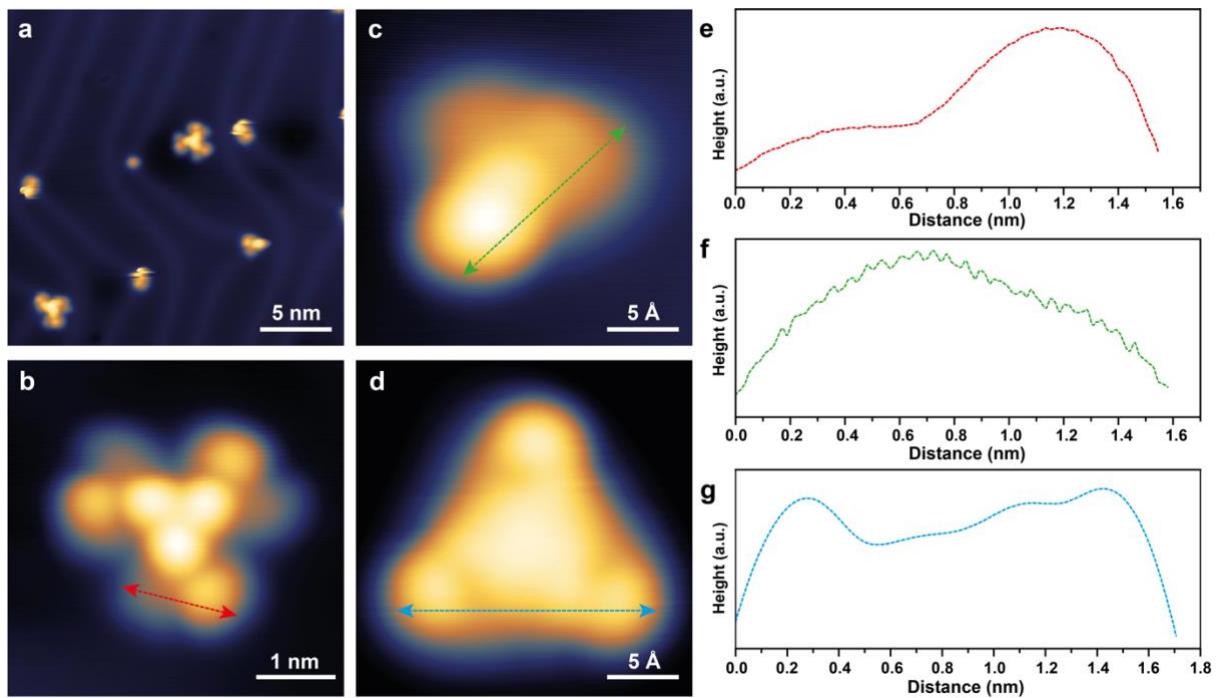


Fig. S10. Structural analysis of TA-H molecule and molecule 1 after mild annealing. (a) Large-scale STM image of the as-deposited TA-H molecule on Au(111) surface. (b) Magnified STM image of a self-assemble TA-H trimmer. (c) Magnified STM image of an isolated TA-H molecule on Au(111) surface. (d) Magnified STM image of an isolated molecule **1** on Au(11) surface. (e-g) Line-profile of the single molecular edge lengths collected from (b-d), respectively. Scanning parameters: (a-c): $V_s = 1.0$ V, $I_t = 50$ pA; (d): $V_s = 1.0$ V, $I_t = 200$ pA. All the STM images are collected by a metal tip.

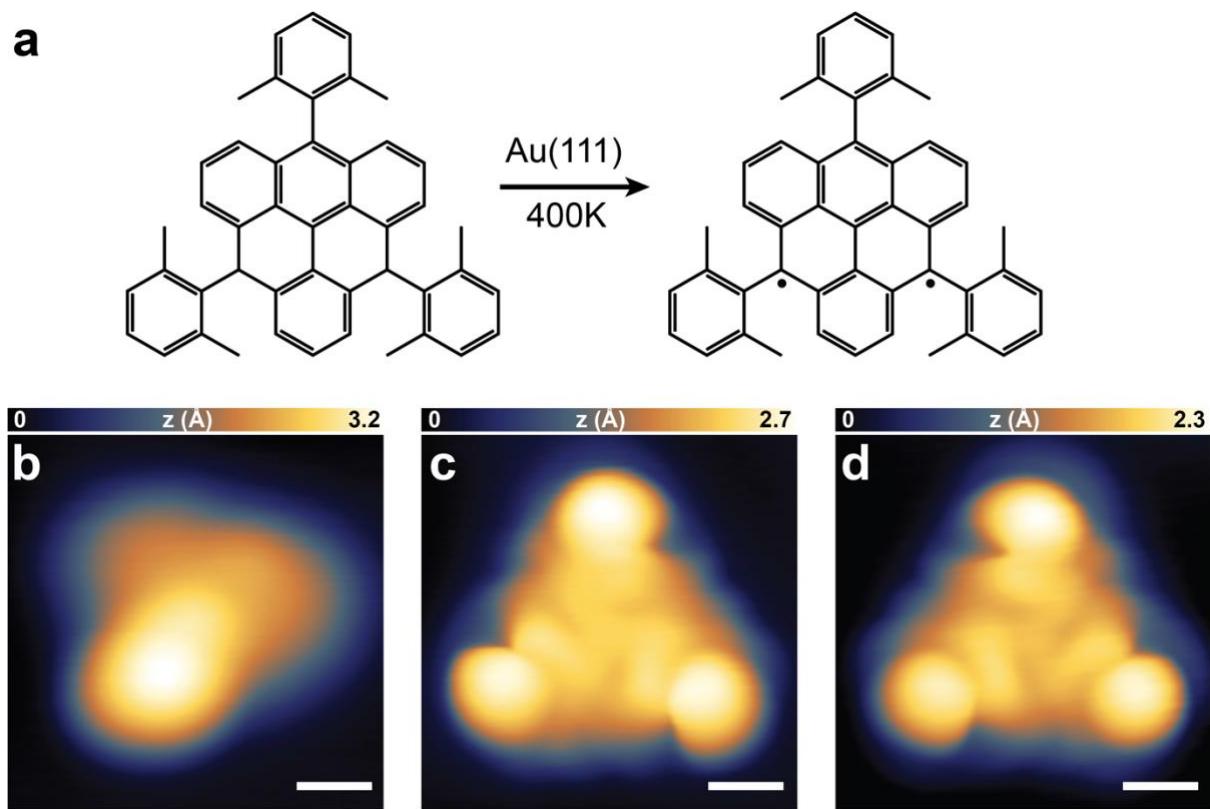


Fig. S11. On-surface formation of molecule 1 derived from molecule TA-H by thermal annealing and STM images of molecule 1 on Au(111) surface with different chiral configurations. (a) Schematic illustration of the synthetic approach from molecule TA-H to molecule 1 on Au(111) by thermal annealing. (b) Magnified STM image of a single TA-H molecule. (c) Clockwise (CW) configuration. (d) Counter-clockwise (CCW) configuration. (c) and (d) are recorded CO functionalized tip. Scale bar: 5 Å. Molecule TA-H was prepared using the procedures reported in the reference (11).

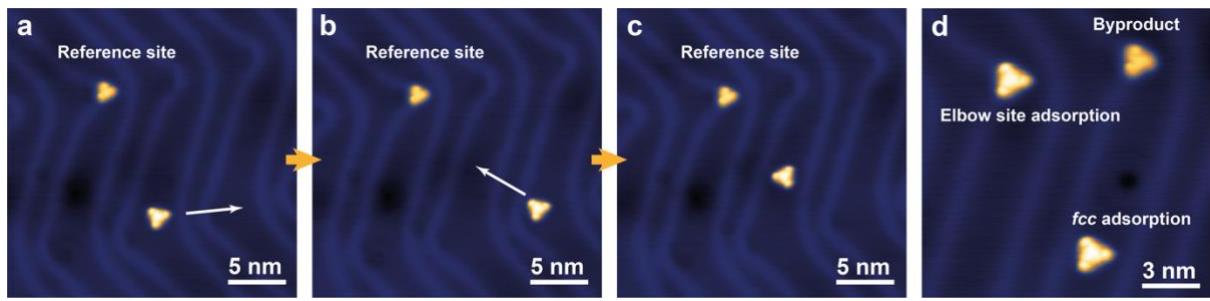


Fig. S12. Molecule 1 with different adsorption configurations on Au(111) surface. (a-c) Lateral movement of a molecule **1** on Au(111) surface *via* atomic manipulation. The white arrows indicate the direction of the movements. The reference site is a byproduct after the annealing on the surface. Note that the topography of molecule **1** does not appear to have obvious changes during the manipulation process. (**d**) STM image of molecule **1** with two typical adsorption configurations: The elbow site of herringbone reconstruction and the *fcc* site of Au(111) surface. The topography of two molecules does not to show obvious difference. Scanning parameters: $V_s = 0.5$ V, $I_t = 400$ pA.

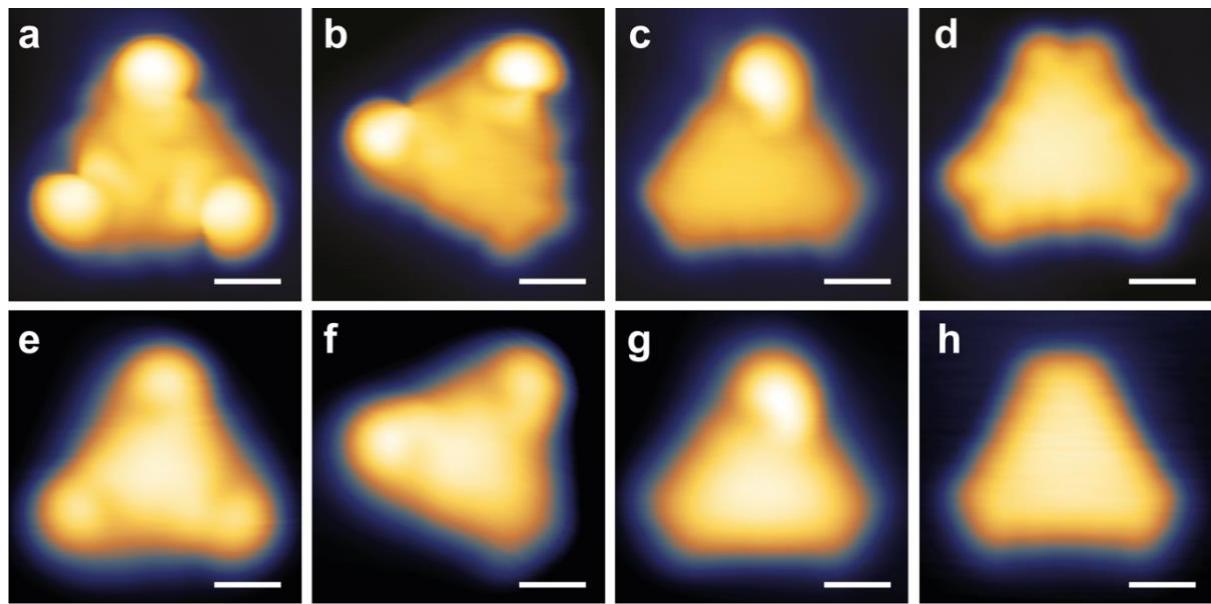


Fig. S13. STM images of molecule 1 to molecule 4 using different tips. (a-d) STM images of molecule **1** to molecule **4** recorded by CO-functionalized tip. (e-h) STM images of molecule **1** to molecule **4** recorded by metal tip. Scanning parameters: $V_s = 0.5$ V, $I_t = 400$ pA. Scale bar: 5 Å.

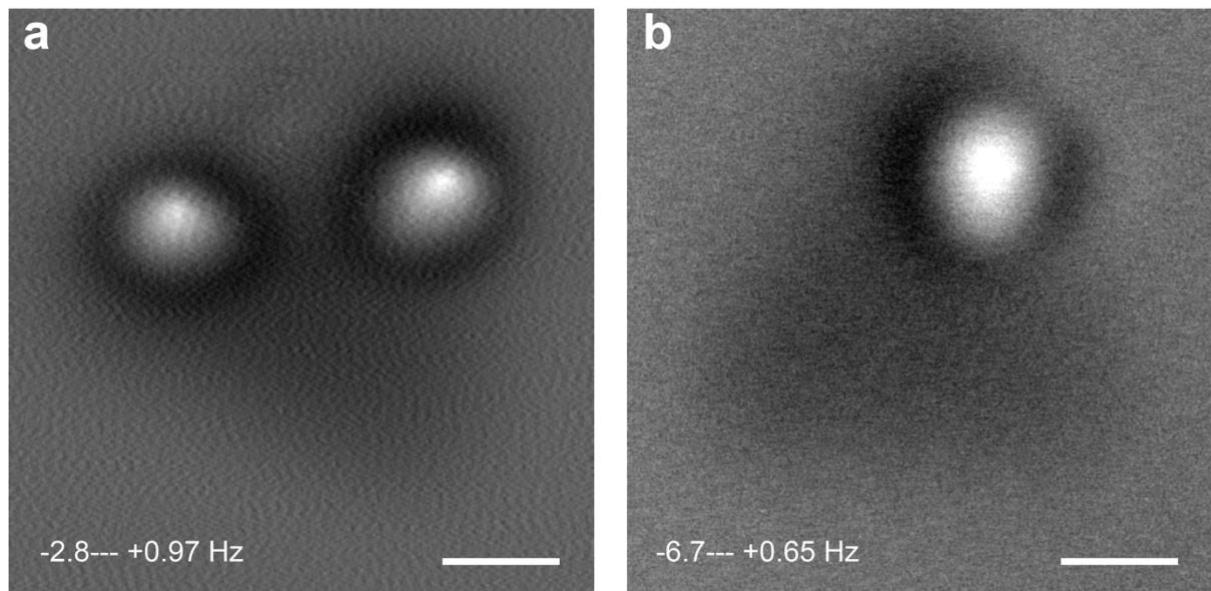


Fig. S14. AFM images of molecule 2 and 3 at large tip-sample distance. Additional AFM images of molecule **2** (**a**) and molecule **3** (**b**) at a large tip-sample distance. Dashed lines indicate the blur cyclized areas that cannot be resolved at a large tip-sample distance. Scanning parameters: setpoint before switching off the feedback: $V_s = 0.5$ V, $I_t = 400$ pA. Scale bar: 4 Å.

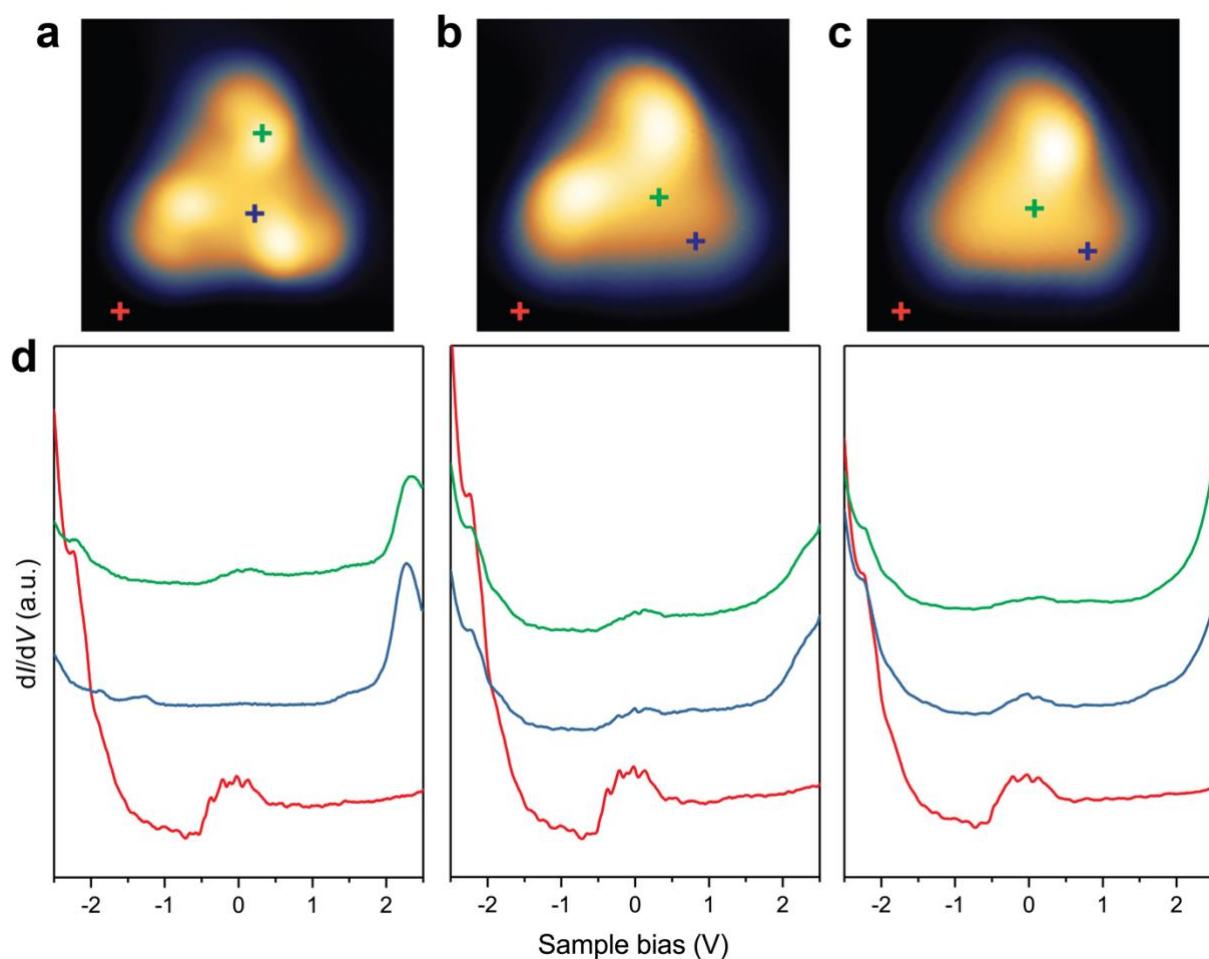


Fig. S15. Large bias range dI/dV spectra of molecule 1, 2 and 3. (a-c) STM images of molecule 1, molecule 2, and molecule 3 on Au(111) and the corresponding dI/dV spectra in a large sample bias window (d). The colours of the curves correspond to the crosses in the STM image, indicating the positions where the dI/dV spectra are recorded. Imaging parameters: $V_s = 2.5$ V, $I_t = 400$ pA.

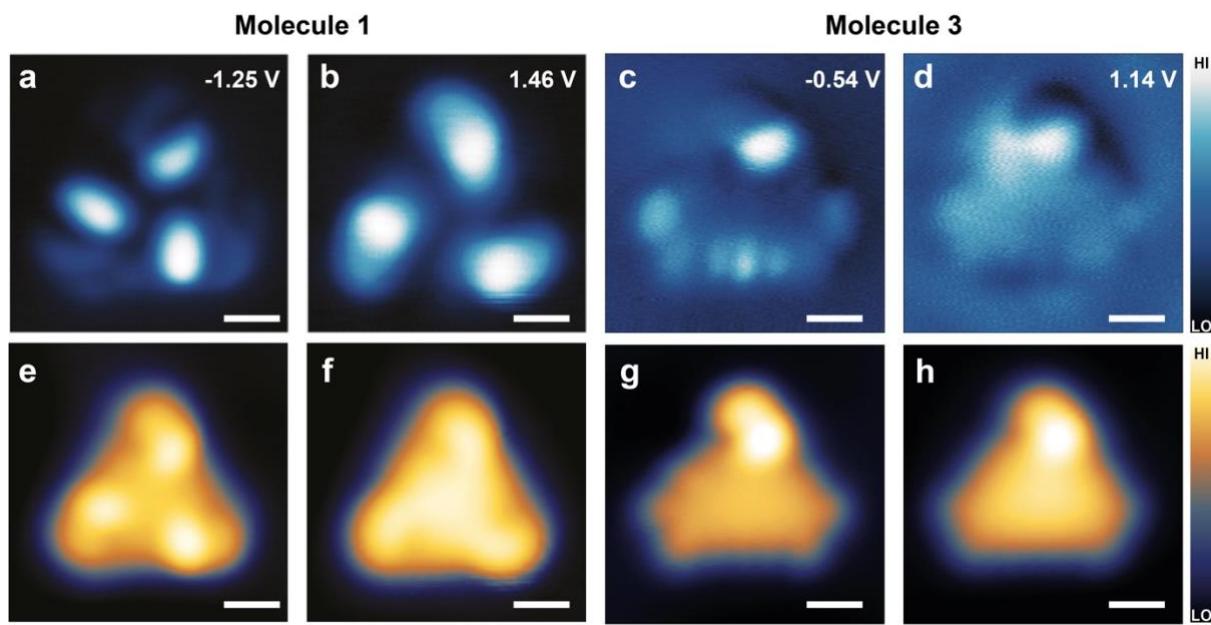


Fig. S16. dI/dV imaging of the frontier molecular orbitals of molecule 1 and molecule 3. Constant-current dI/dV maps of the SOMOs and SUMOs of molecule 1 (**a-b**) and the corresponding STM images (**e-f**) taken simultaneously. Constant-current dI/dV maps of the SOMOs and SUMOs of molecule 3 (**c-d**) and the corresponding STM images (**g-h**) taken simultaneously.

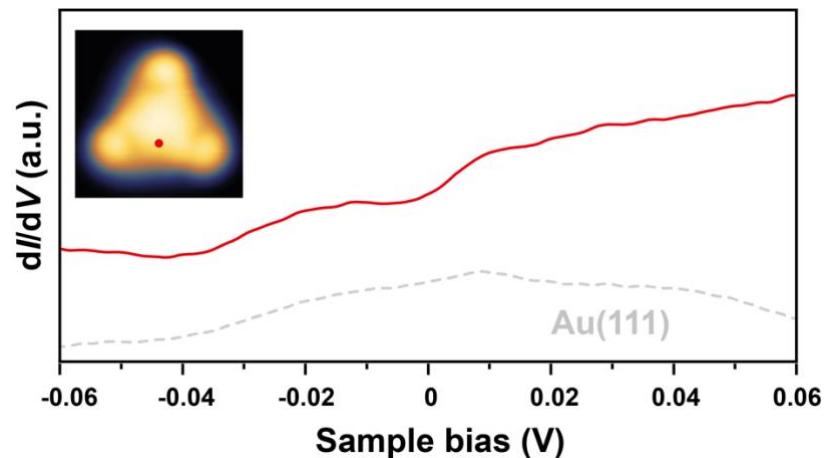


Fig. S17. dI/dV spectra recorded over a molecule **1** near the E_F . Setpoint: $V_s = 0.06$ V, $I_t = 1.2$ nA. Modulation voltage: 2 mV.

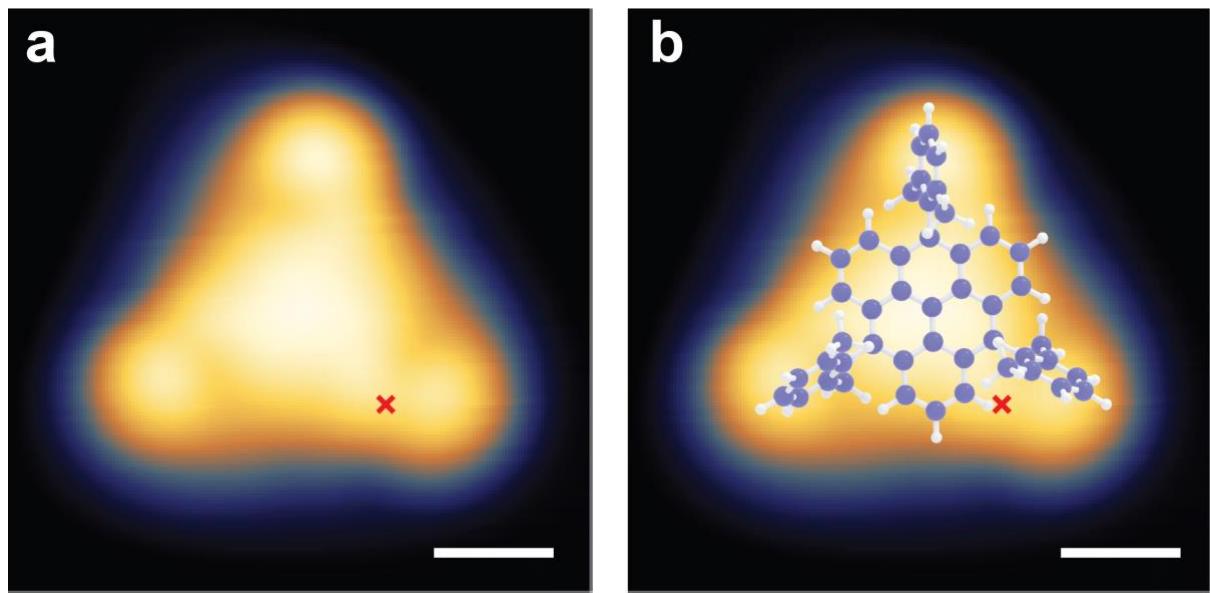


Fig. S18. Tip position of a successful cyclodehydrogenation process. The red cross marks the position of the tip where the voltage pulse is applied. Scale bar: 5 Å. The purple and white balls represent carbon and hydrogen atoms, respectively.

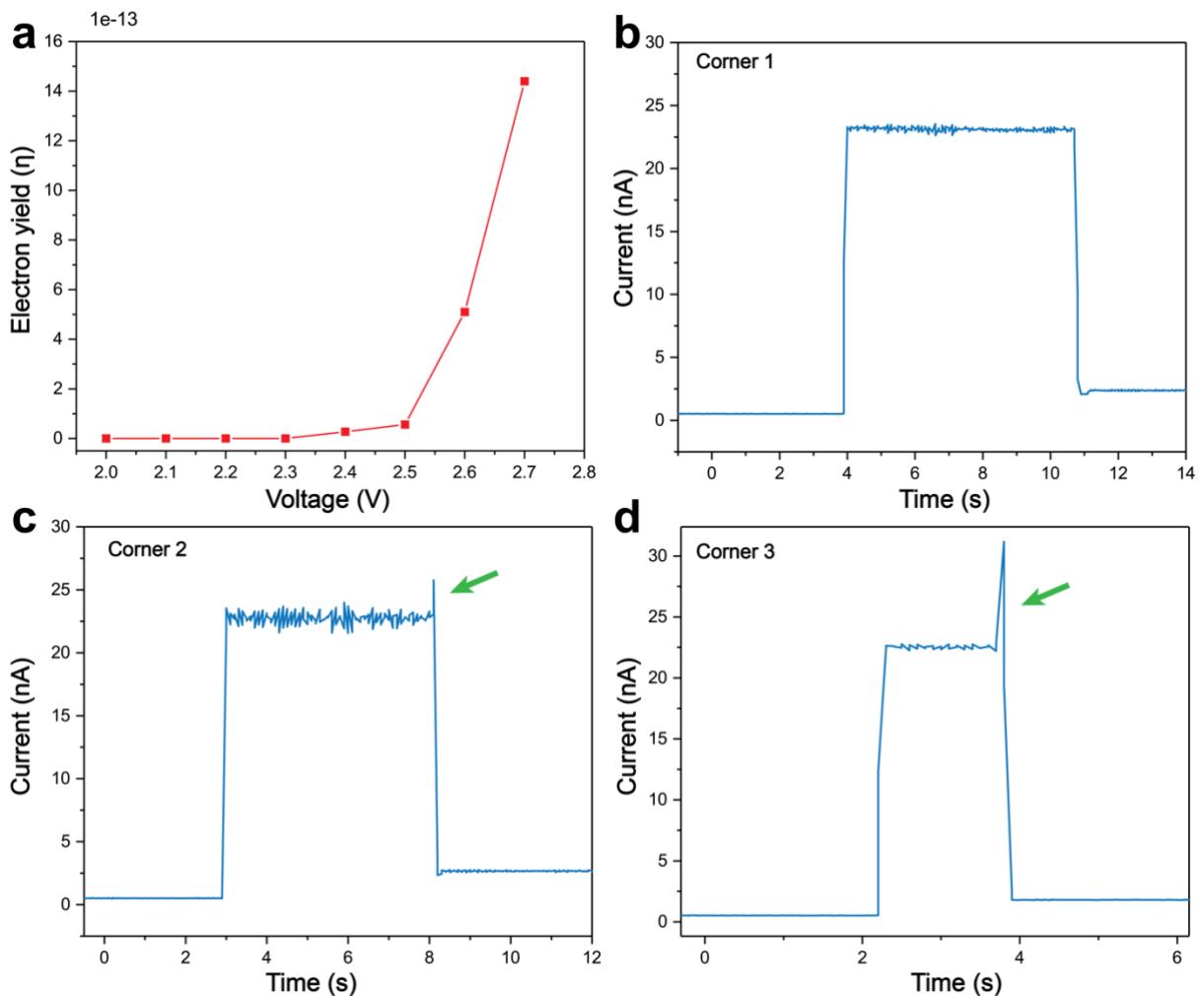


Fig. S19. Reaction details of the cyclodehydrogenation. (a) Voltage threshold of triggering a cyclodehydrogenation of a molecule **1** on Au(111) surface. (b) I/t curve of the cyclization of molecule **1**. (c) I/t curve of the cyclization of molecule **2**. (d) I/t curve of the cyclization of molecule **3**.

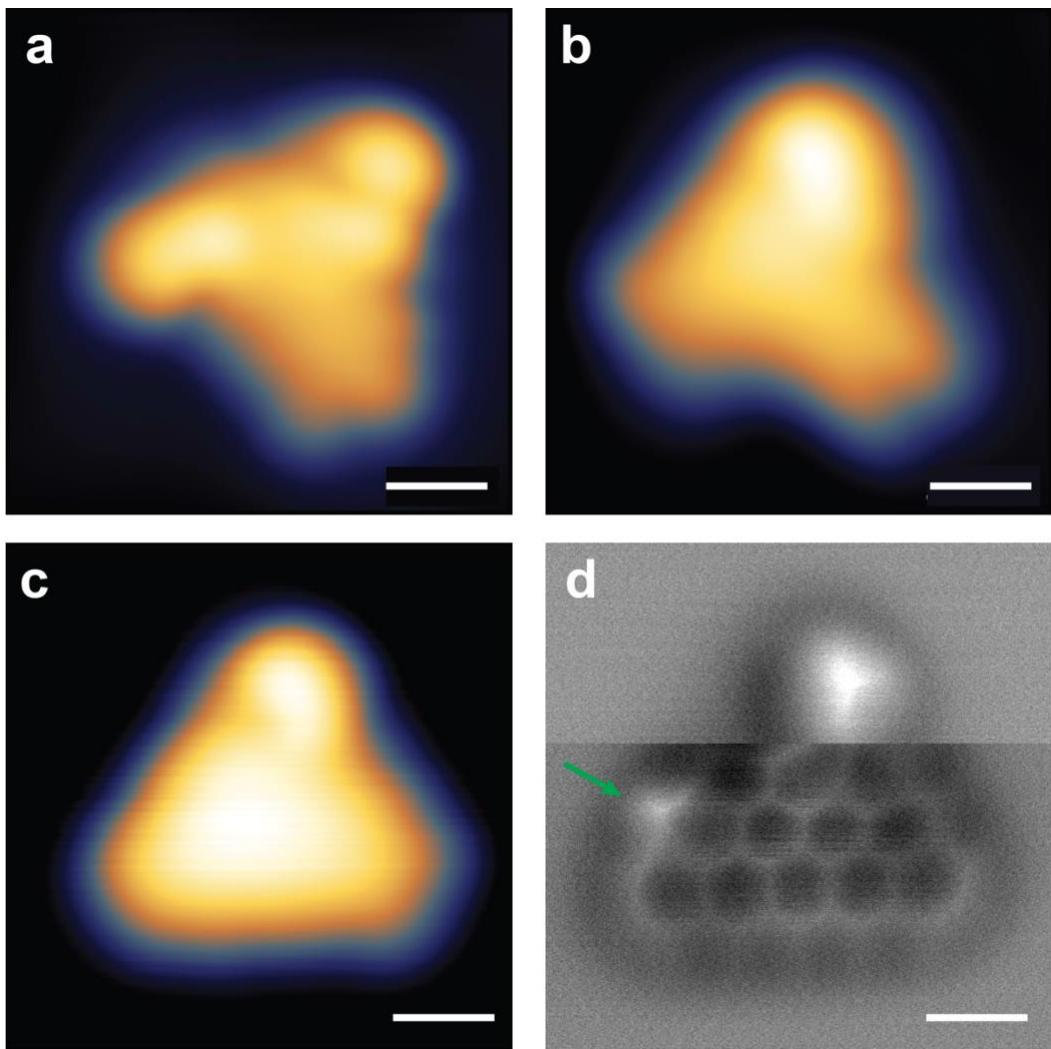


Fig. S20. STM and AFM images of non-fully cyclized products. (a-b) STM images of examples of unsuccessful cyclization during the first (a) and the second (b) cyclization process. (c) STM image of a cyclized molecule **3** with a CH₂ site on the left edge, showing a slightly different corner feature compared with the fully-cyclodehydrogenated one. (d) The corresponding AFM image of the same CH₂-molecule **3** in (c). The -CH₂ site on the left edge is indicated by the green arrow. The tip-sample distance was increased for ~200 pm at the boundary between the planar region and the *m*-Xylene moiety. Scale bar: (a-c) 5 Å, (d) 4 Å.

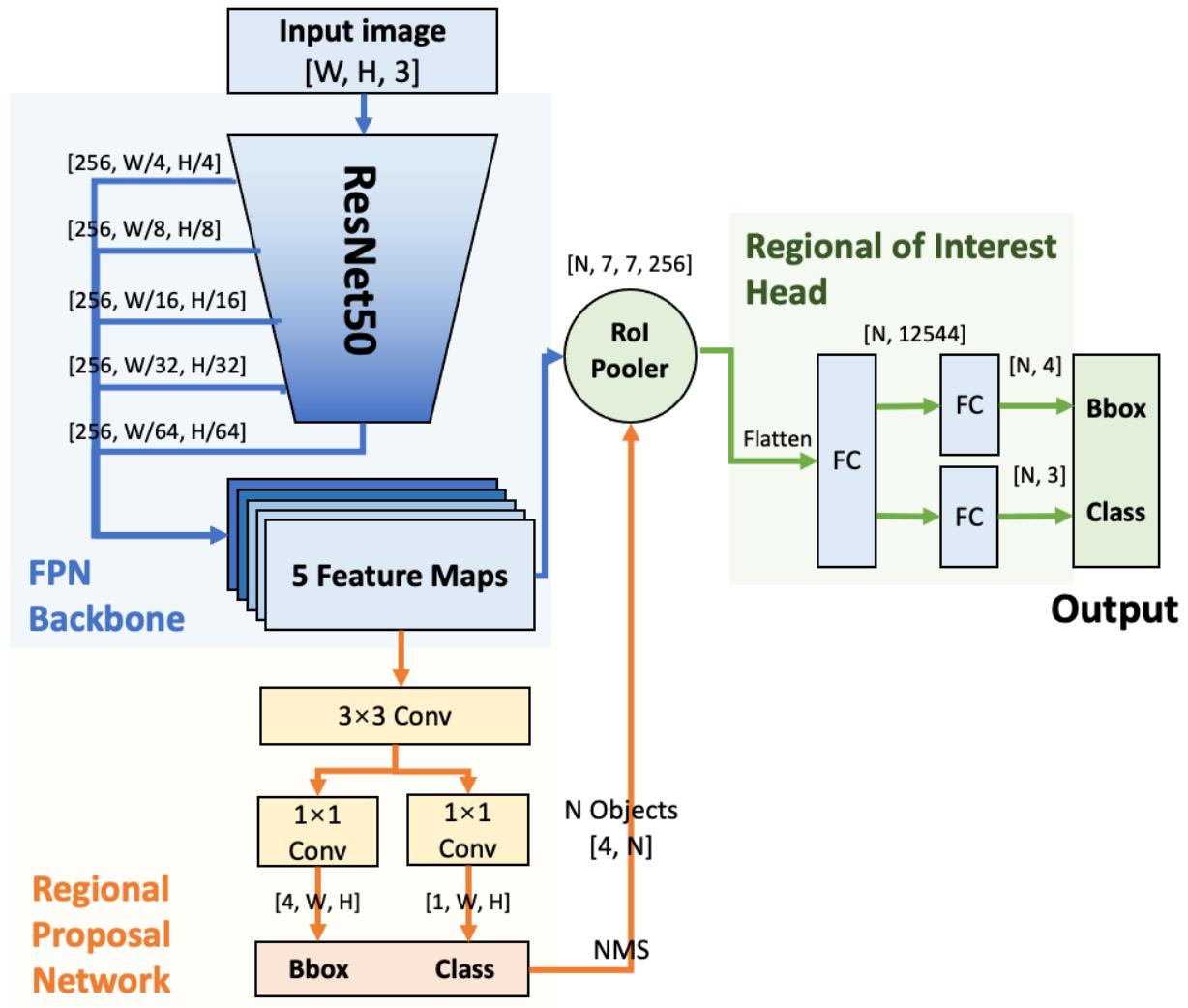


Fig. S21. Block diagram of Faster R-CNN (TI module).

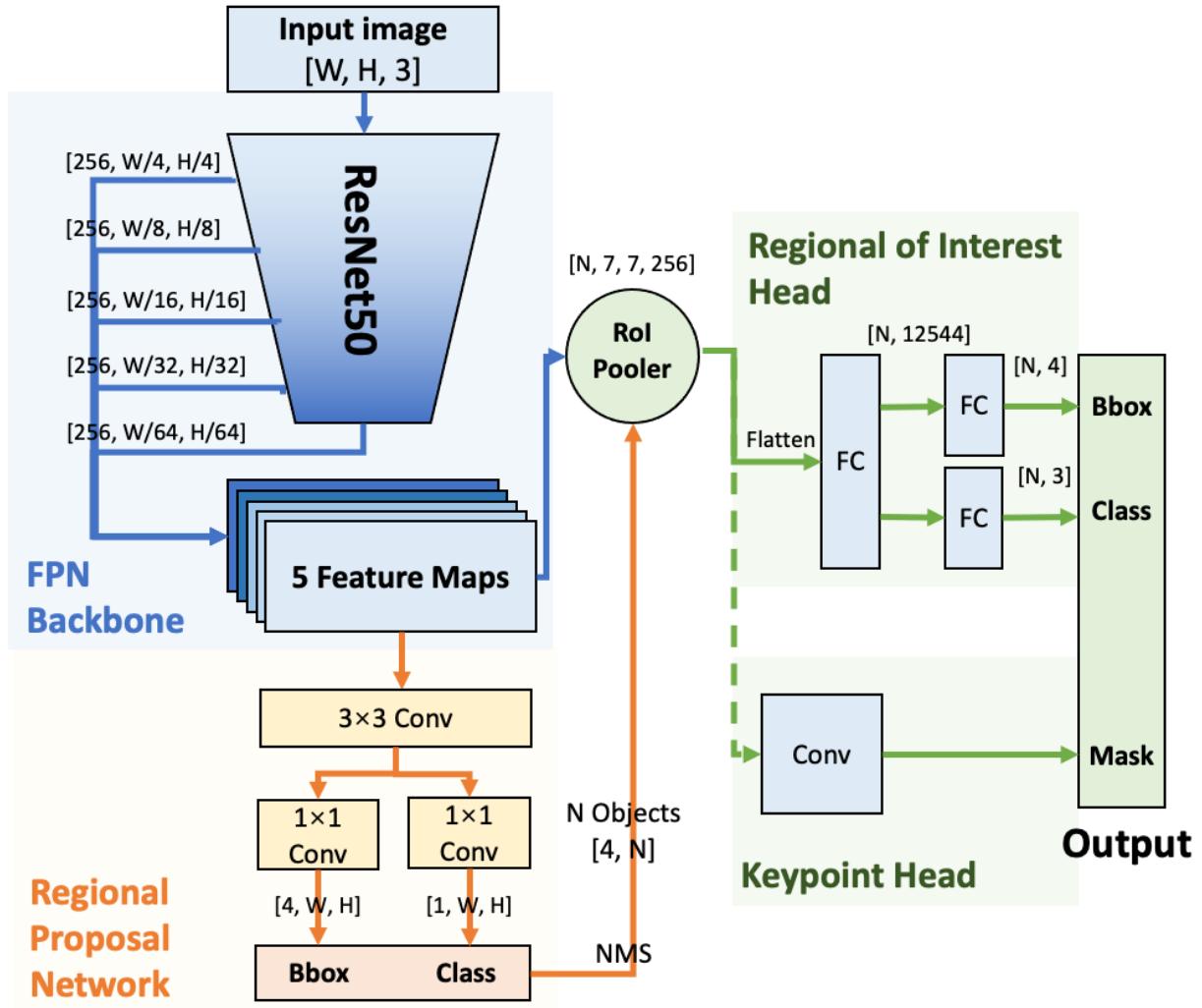


Fig. S22. Block diagram of key point R-CNN (SJ module).

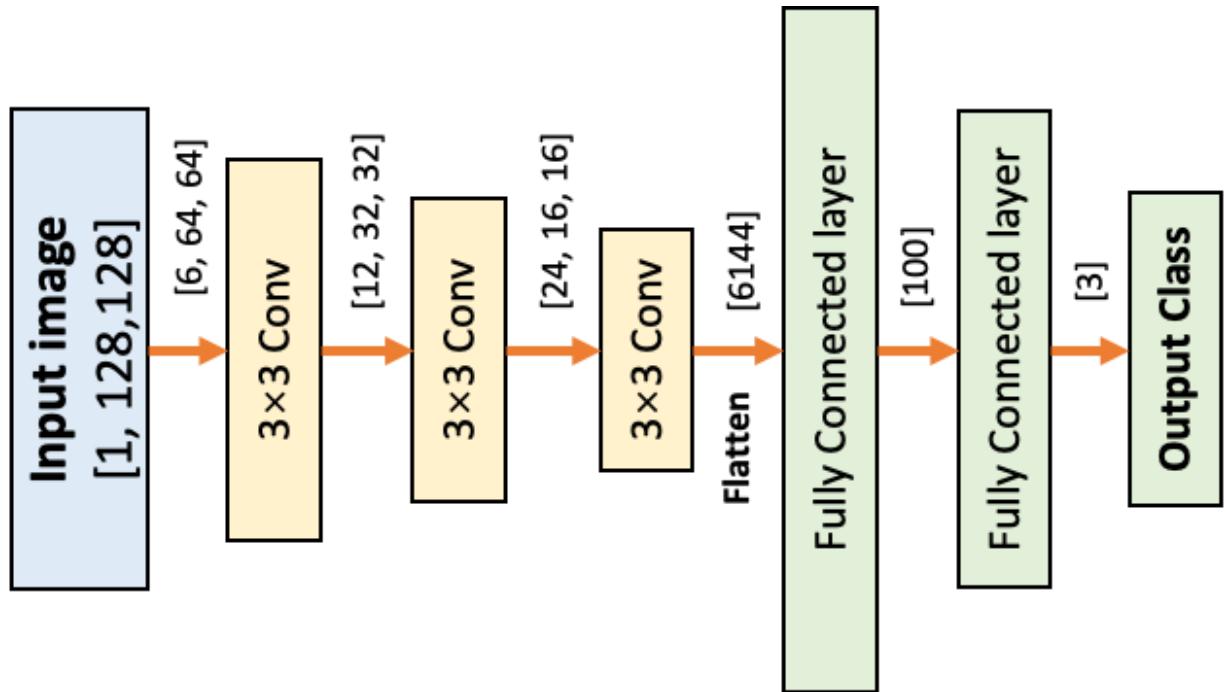


Fig. S23. Block diagram of product classification CNN (PC module).

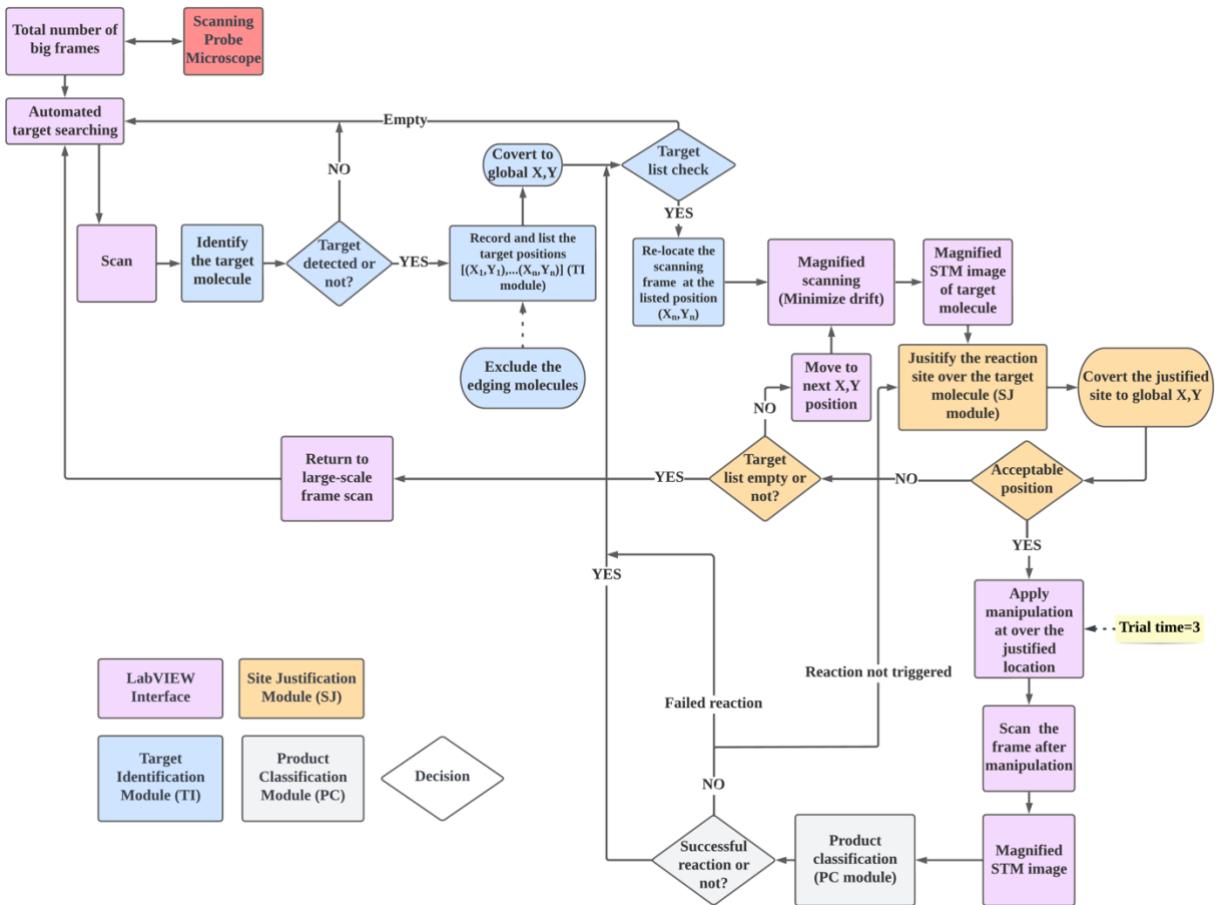


Fig. S24. Overall design of the STM-AI system.

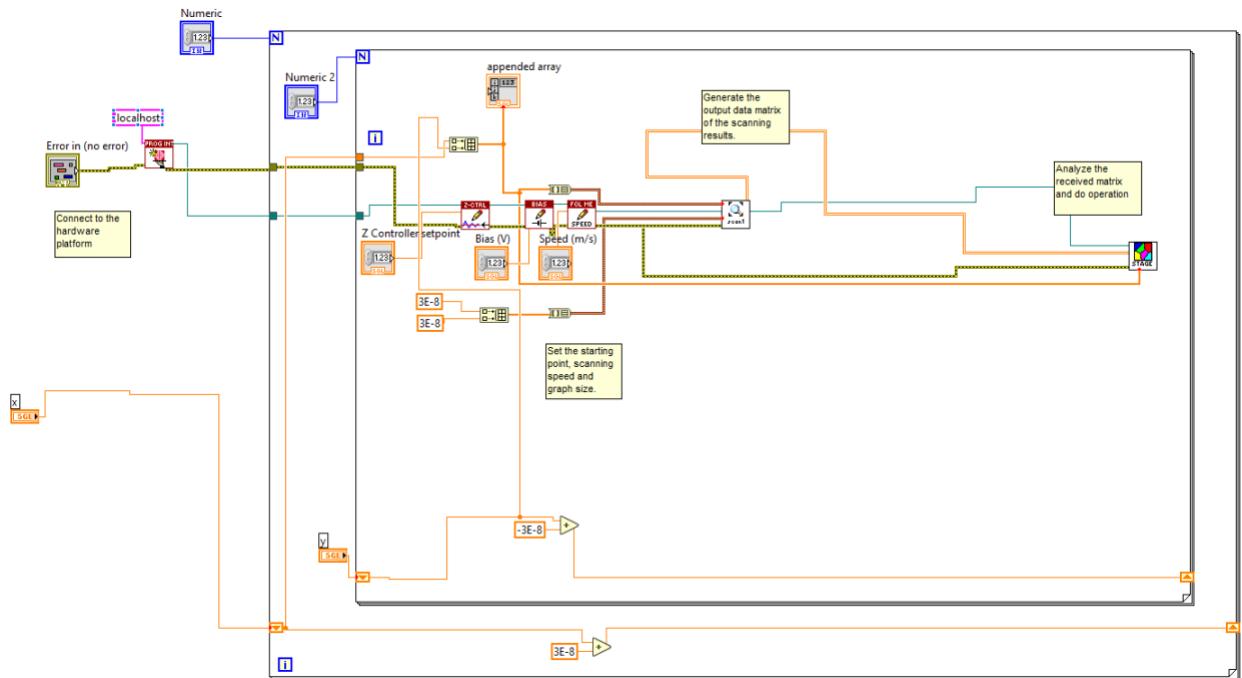


Fig. S25. The main structure of the STM-AI interface.

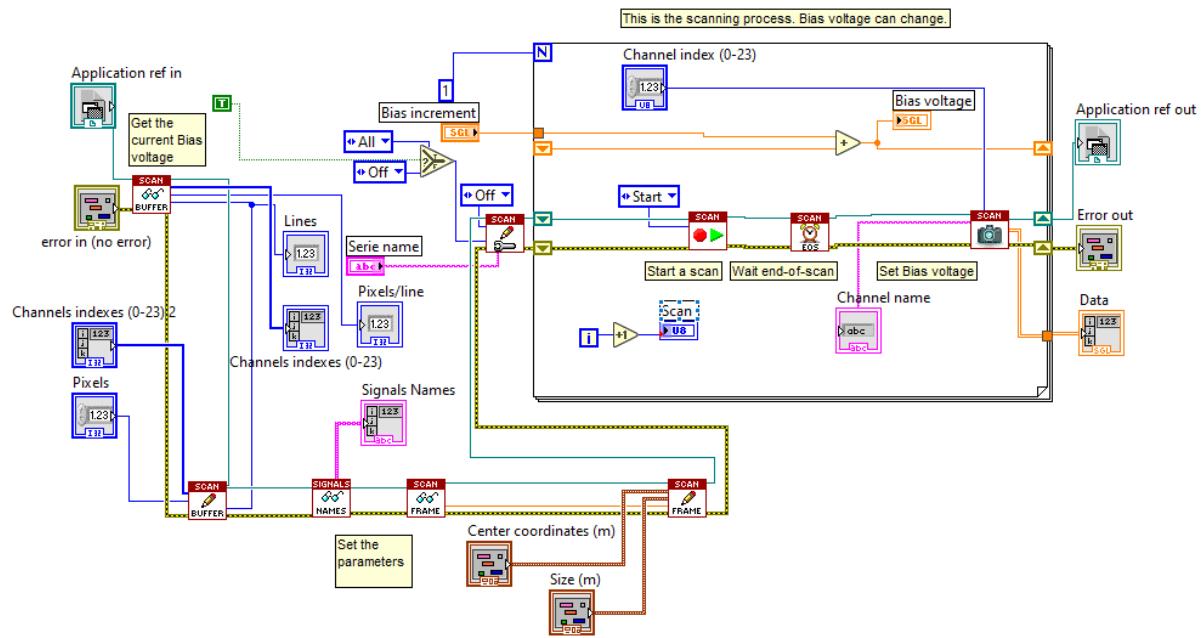


Fig. S26. The structure of the automated scanning module in the STM-AI interface.

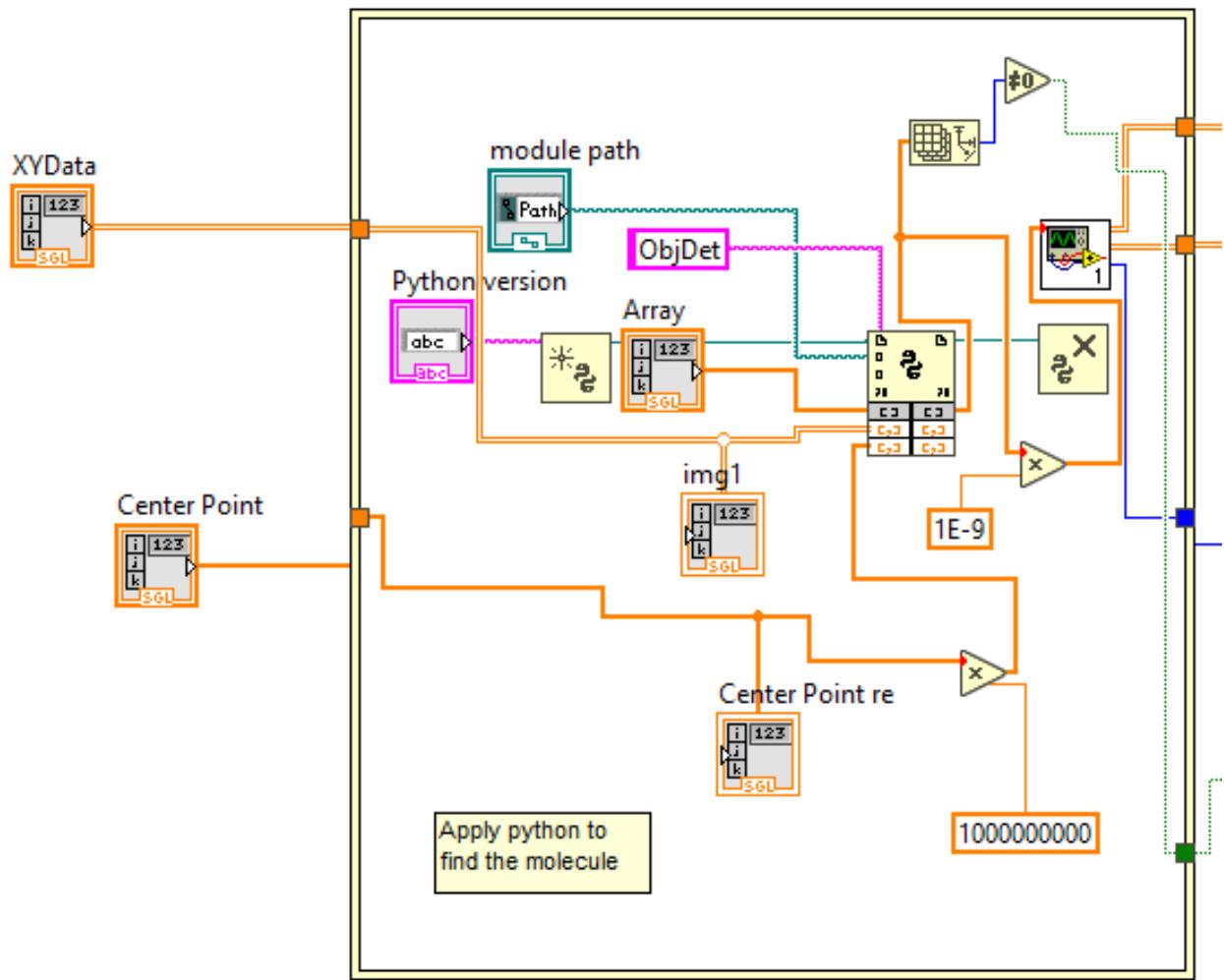


Fig. S27. The sub-structure of STM-AI interface which is responsible for communicating the TI module.

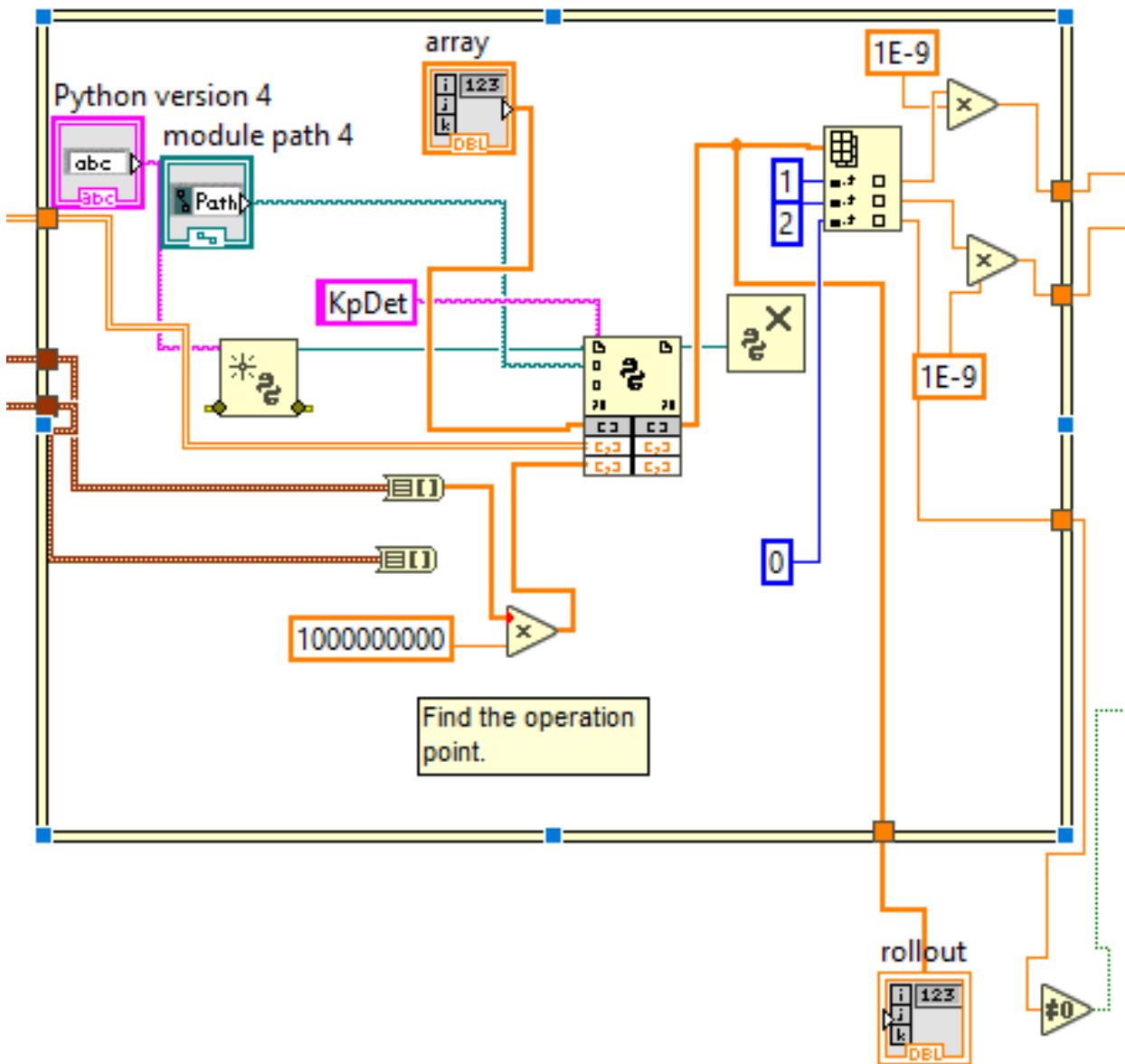


Fig. S28. The sub-structure of STM-AI interface which is responsible for communicating the SJ module.

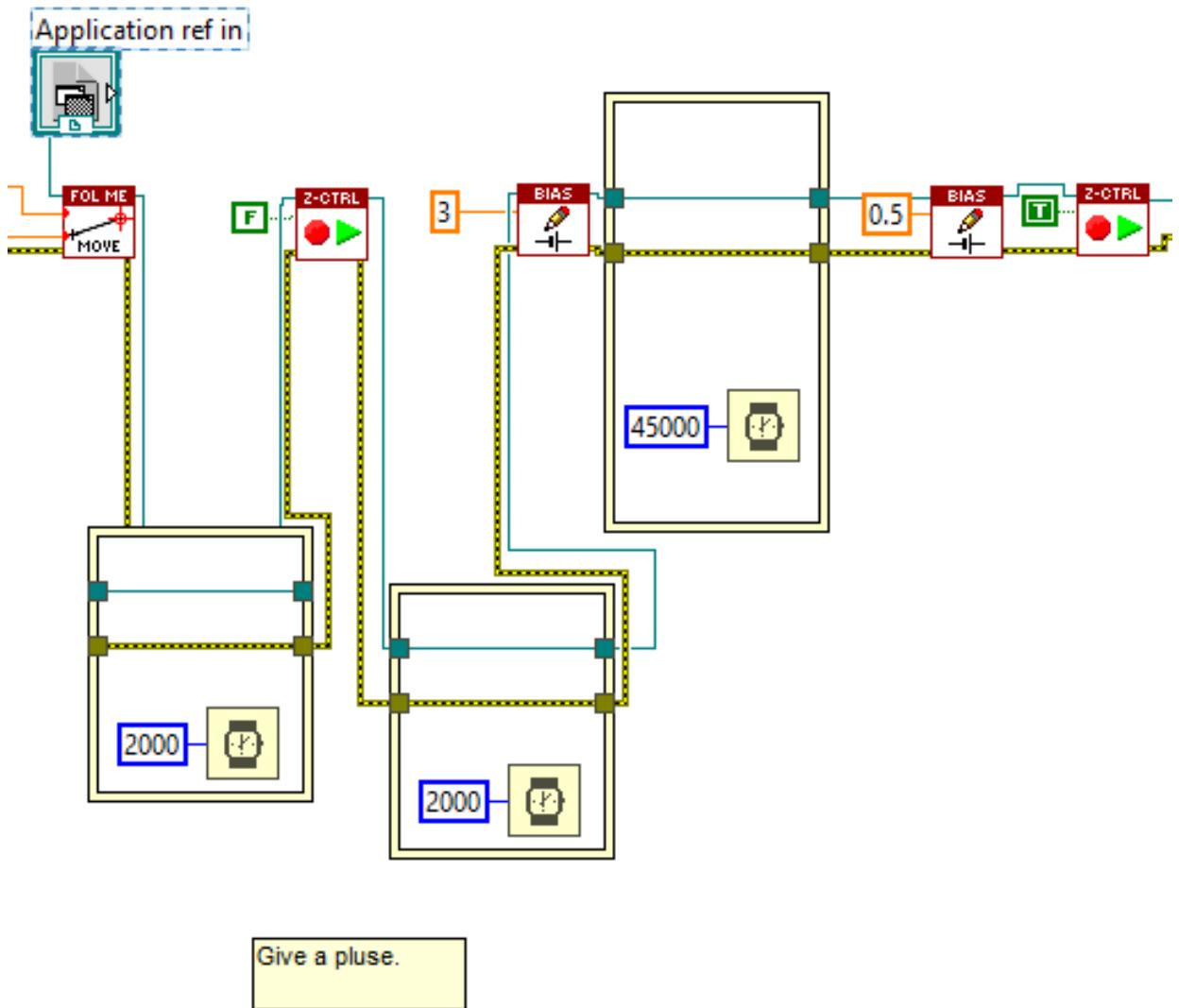


Fig. S29. The sub-structure of STM-AI interface which is responsible for conducting the manipulation.

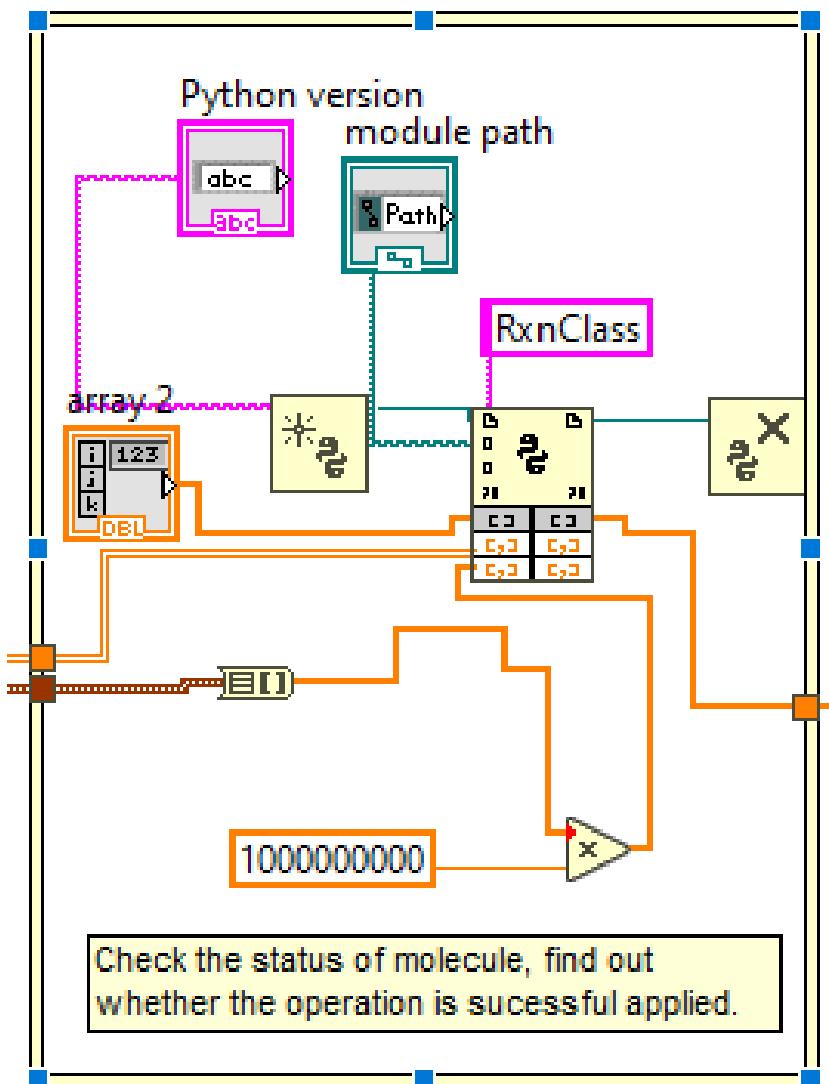


Fig. S30. The sub-structure of STM-AI interface which is responsible for communicating the PC module.

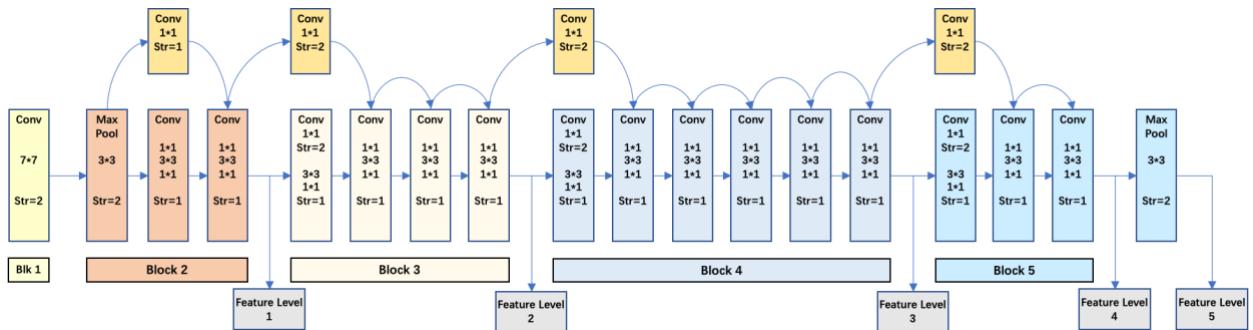


Fig. S31. ResNet 50 architecture.

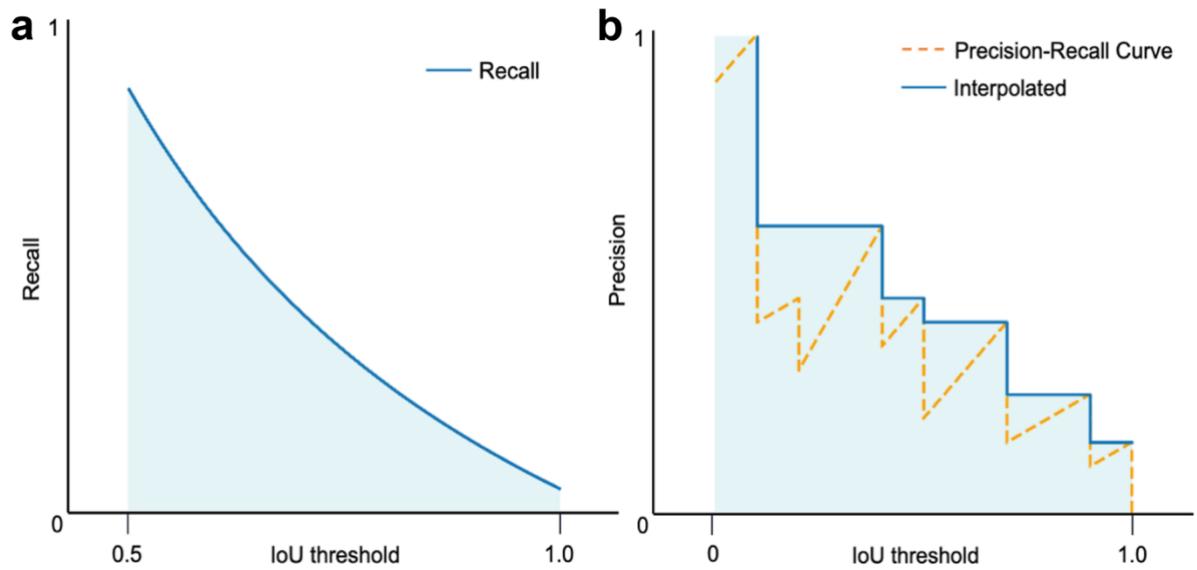


Fig. S32. Schematic Recall-IoU curve and Precision-Recall curve for general explanation purpose.

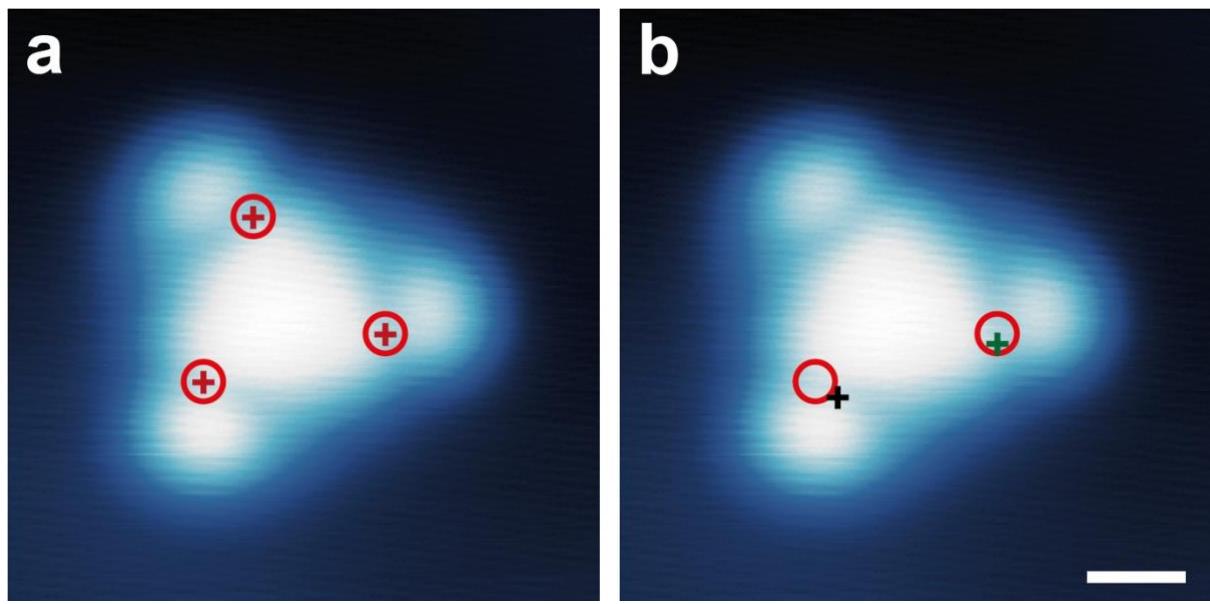


Fig. S33. Reaction site recognition rate (RSRR) justification. (a) The STM image of a target molecule **1** with proposed reference reaction sites given by the expert knowledge. The red crosses represent the reference sites, and the red circles represent the effective regions. (b) two examples of the predicted reaction sites. The green cross locates within the circle (correct), while the black cross locates close to but outside the circle (incorrect). Scale bar: 5 Å.

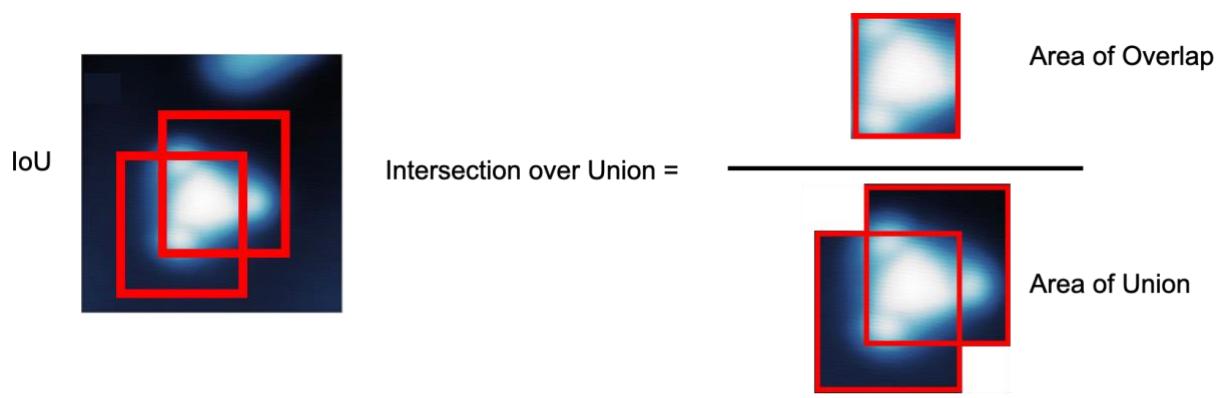


Fig. S34. IoU visualization diagram.

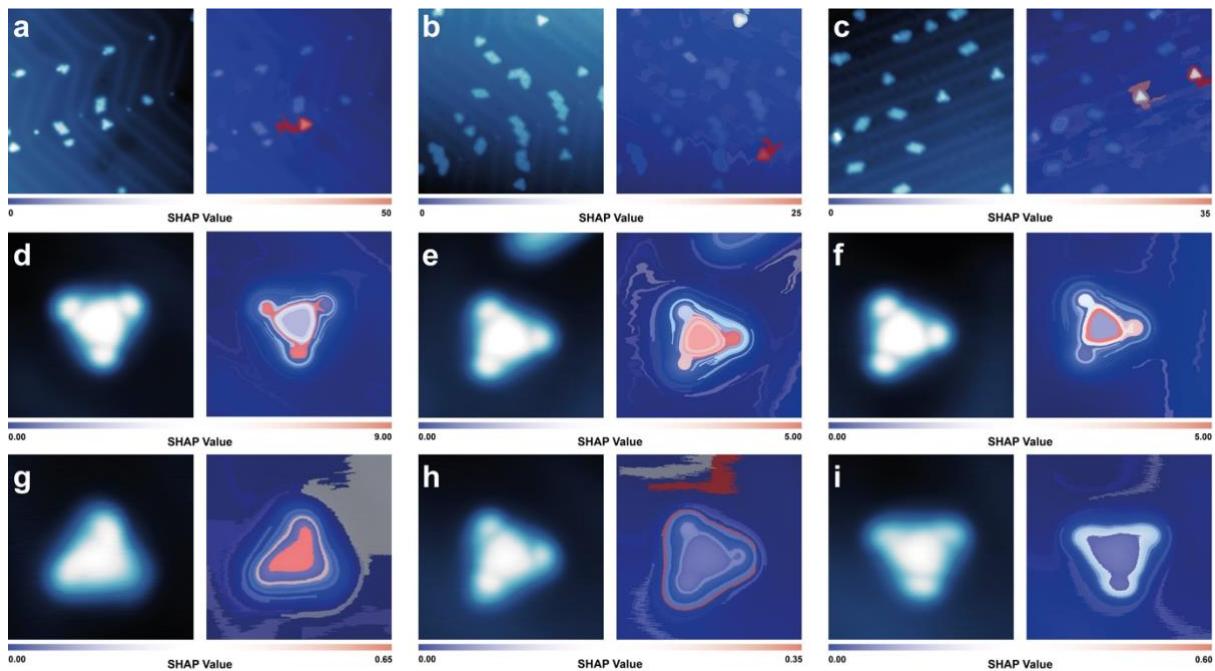


Fig. S35. Examples of SHAP analysis of three deep neural modules. (a-c) SHAP analysis of TI modules in large-scale STM images. (d-f) SHAP analysis of SJ modules in magnified STM images of molecule **1**. (g-i) SHAP analysis of PC modules in magnified STM images molecule **2** (g) and **1** (h-i). For each panel, the left image is the original STM image and the right one is SHAP analysis results.

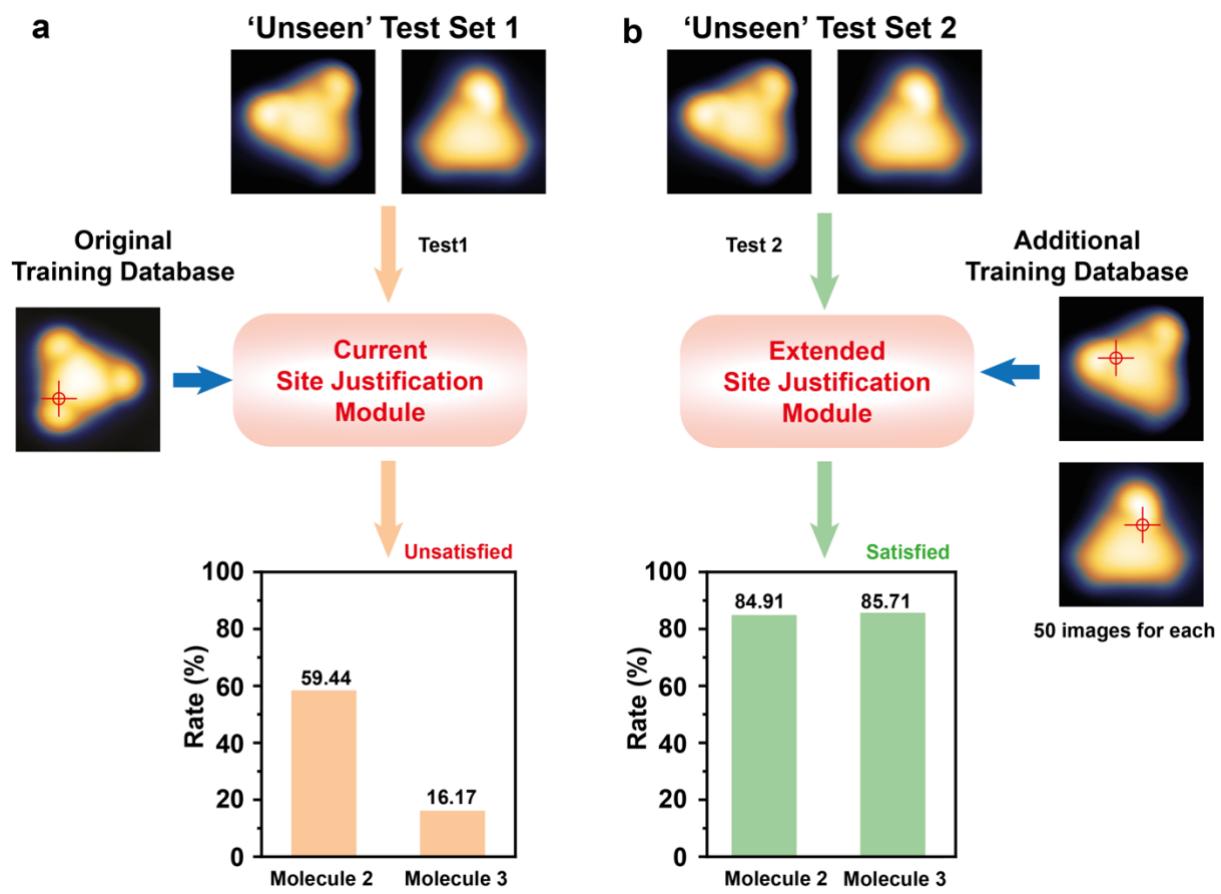
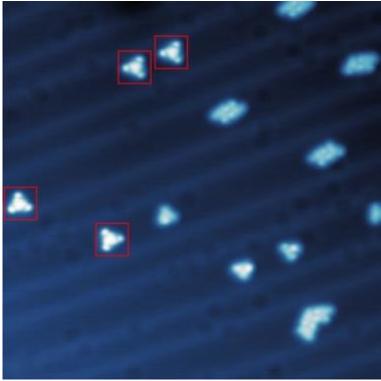
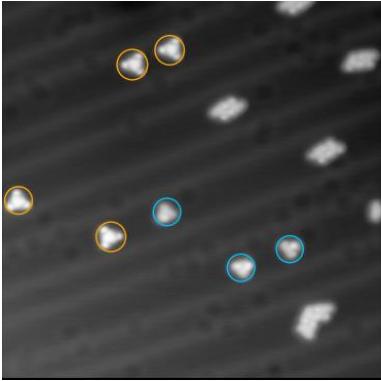
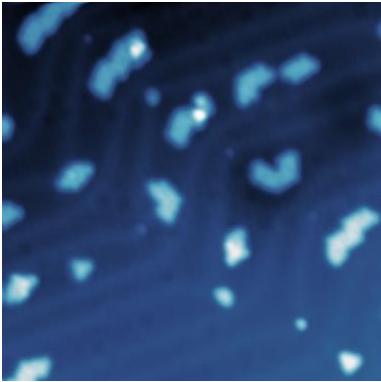
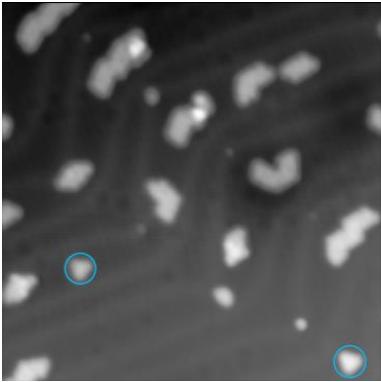
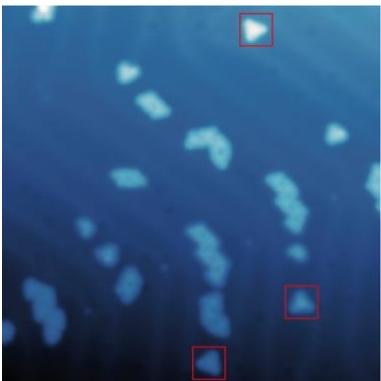
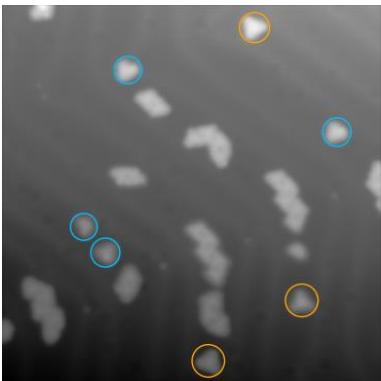


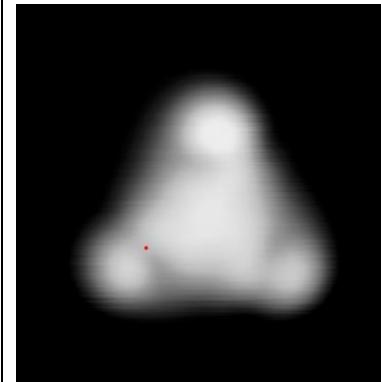
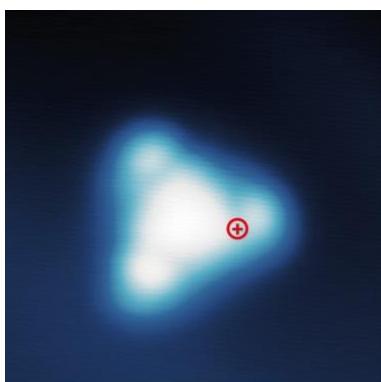
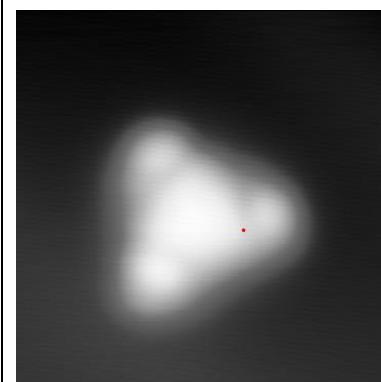
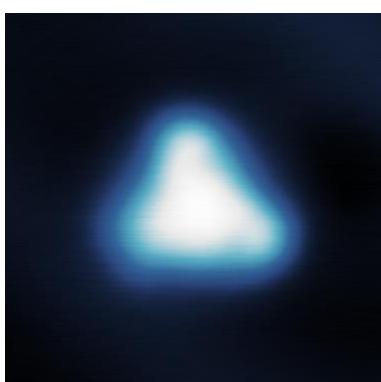
Fig. S36. Statistics of the RSRR (reaction site recognition rate) performance of different SJ modules embedded in the CARP framework. (a) The orange columns represent the RSRR value of justifying the possible reaction sites over molecule **2** and molecule **3** using the current SJ module (specifically trained for molecule **1**), respectively. (b) The green columns represent the RSRR results of justifying the possible reaction sites over molecule **2** and molecule **3** using the extended-SJ module (with limited volume database training), respectively.

Table S1. Total system energies of triplet state of molecule **1** and molecule **3** with respect to that of the singlet closed-shell state.

Energy (eV)	Molecule 1 /Au(111)	Molecule 3 /Au(111)
Triplet (spin)	-0.14	-0.049
Singlet (no spin)	0	0

Table S2. Hand-labelled STM images and target-predicted STM images of different molecules

No.	Hand-labelled STM images	Prediction STM images	Description
1			<p>Large-scale STM image with target molecule (Clear contrast). Red rectangle: Target molecule labelled by hand. Yellow circle: Target molecule identified by the prediction model. Blue circle: Molecule with similar shape but not the target, classified by model.</p>
2			<p>Large-scale STM image without target molecule (Clear contrast). Blue circle: Molecule with similar shape but not the target, classified by model.</p>
3			<p>Large-scale STM image with target molecule (Blurred contrast). Red rectangle: Target molecule labelled by hand. Yellow circle: Target molecule identified by the prediction model. Blue circle: Molecule with similar shape but not the target, classified by model.</p>

4			<p>Magnified STM image of target molecule (Clear contrast). Red cross: Possible reaction site labelled by hand. Red circle: Effective region of reaction site prediction. Red dot: Reaction site predicted by model.</p>
5			<p>Magnified STM image of target molecule (Blurred contrast). Red cross: Possible reaction site labelled by hand. Red circle: Effective region of reaction site prediction. Red dot: Reaction site predicted by model.</p>
6		NA	<p>Magnified STM image of product molecule.</p>

Supplementary References

- 63 Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558-561 (1993). <https://doi.org/10.1103/physrevb.47.558>
- 64 Kresse, G. & Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169-11186 (1996). <https://doi.org/10.1103/physrevb.54.11169>
- 65 Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865-3868 (1996). <https://doi.org/10.1103/PhysRevLett.77.3865>
- 66 Blochl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953-17979 (1994). <https://doi.org/10.1103/physrevb.50.17953>
- 67 Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32**, 1456-1465 (2011). <https://doi.org/10.1002/jcc.21759>
- 68 Toyoda, K., Nozawa, K., Matsukawa, N. & Yoshii, S. Density Functional Theoretical Study of Graphene on Transition-Metal Surfaces: The Role of Metal d-Band in the Potential-Energy Surface. *J. Phys. Chem. C* **117**, 8156-8160 (2013). <https://doi.org/10.1021/jp311741h>
- 69 Too, H. L., Guo, N., Zhang, C. & Wang, Z. Importance of Sugar-Phosphate Backbone and Counterions to First-Principles Modeling of Nucleobases. *J. Phys. Chem. B* **126**, 5744-5751 (2022). <https://doi.org/10.1021/acs.jpcb.2c02621>
- 70 Nguyen, G. D. *et al.* Atomically precise graphene nanoribbon heterojunctions from a single molecular precursor. *Nat. Nanotechnol.* **12**, 1077-1082 (2017). <https://doi.org/10.1038/nnano.2017.155>
- 71 Paier, J., Marsman, M. & Kresse, G. Why does the B3LYP hybrid functional fail for metals? *J. Chem. Phys.* **127**, 024103 (2007). <https://doi.org/10.1063/1.2747249>
- 72 Momma, K. & Izumi, F. VESTA 3for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Crystallogr.* **44**, 1272-1276 (2011). <https://doi.org/10.1107/s0021889811038970>
- 73 Wang, V., Xu, N., Liu, J.-C., Tang, G. & Geng, W.-T. VASPKIT: A user-friendly interface facilitating high-throughput computing and analysis using VASP code. *Comput. Phys. Commun.* **267** (2021). <https://doi.org/10.1016/j.cpc.2021.108033>
- 74 Kouwenhoven, L. & Glazman, L. Revival of the Kondo effect. *Phys. World* **14**, 33-38 (2001). <https://doi.org/10.1088/2058-7058/14/1/28>
- 75 Wahl, P. *et al.* Kondo temperature of magnetic impurities at surfaces. *Phys. Rev. Lett.* **93**, 176603 (2004). <https://doi.org/10.1103/PhysRevLett.93.176603>
- 76 Ternes, M., Heinrich, A. J. & Schneider, W. D. Spectroscopic manifestations of the Kondo effect on single adatoms. *J. Phys. Condens. Matter* **21**, 053001 (2009). <https://doi.org/10.1088/0953-8984/21/5/053001>
- 77 Turco, E. *et al.* Observation of the Magnetic Ground State of the Two Smallest Triangular Nanographenes. *JACS Au* **3**, 1358-1364 (2023). <https://doi.org/10.1021/jacsau.2c00666>
- 78 Schneider, M. A. *et al.* Kondo state of Co impurities at noble metal surfaces. *Appl. Phys. A* **80**, 937-941 (2005). <https://doi.org/10.1007/s00339-004-3119-7>
- 79 Pavlicek, N. *et al.* Polyyne formation via skeletal rearrangement induced by atomic manipulation. *Nat. Chem.* **10**, 853-858 (2018). <https://doi.org/10.1038/s41557-018-0067-y>
- 80 He, K., Zhang, X., Ren, S. & Sun, J. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770-778 (2016).