Revealing conductance variation of molecular junctions based on an unsupervised data analysis approach

Shuhui Tao[a,b], Qian Zhang[a,b], Sylvain Pitie[c], Chenguang Liu[d], Yinqi Fan[a,b], Chun Zhao[d], Mahamadou Seydou[c], Yannick J. Dappe[e], Richard J. Nichols[b], and Li Yang [a,b,*]

[a]Department of Chemistry, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China.

[b]Department of Chemistry, University of Liverpool, Liverpool, L69 7ZD, UK.

[c]Université de Paris, ITODYS, CNRS, F-75006 Paris, France.

[d]Department of Electrical and Electronic Engineering, Xi'an-Jiaotong Liverpool University, Suzhou, 215123, China.

[e]SPEC, CEA, CNRS, Université Paris-Saclay, CEA Saclay 91191 Gif-sur-Yvette Cedex, France.

[*]**Corresponding Author:** li.yang@xjtlu.edu.cn

Abstract

For single-molecule electrical junction conductance measurements the data analysis can be a key challenge because of the large datasets and large stochastic variation in junction conformation from one junction formation cycle to the next. Here, a data sorting algorithm was used to analyze the most dominant conductance groups for four molecular systems. The algorithm was applied to data sorting of molecular junctions repetitively formed during STM current-distance (*I*-*s*) retraction traces. The data was recorded using the STM *I*(*s*) technique to form molecular junctions, which avoids direct contact between the gold STM tip and gold substrate. The algorithm has been employed to identify different dominant conductance features for two saturated molecular bridges and two conjugated molecular bridges. The former molecular targets featured thiolmethyl and isothiocyanate anchoring groups on each respective end of the polymethylene molecular bridge, while the conjugated molecular wires featured thiol anchoring groups. In each case the sorting algorithm is able to identify two main junction conductance groups related to differing molecular junction configurations.

This algorithm is demonstrated to effectively sort and rationalize complex molecular junction formation datasets, where junction formation probably can be relatively low, and thereby shows potential to quantify hidden features in single-molecule charge transport data.

Introduction

Characterizing, quantifying and understanding electron transfer through molecular junctions is a key challenge in molecular electronics [1–4]. Single molecule measurements based on repetitive junction formation and breaking cycles contains a plethora of physical and chemical information, which could be categorized and further analyzed as mathematic datasets. Investigating such datasets with hand-selection methods can be very challenging, especially if knowledge about the data is limited and *a priori* assumptions about expected data characteristics are to be avoided [5].

From the experimental point of view, the most commonly used approaches for molecular conductance measurements rely on repetitively building stable metal|molecule|metal junctions by collecting the current-distance profiles during junction formation, typically with either scanning tunneling microscope (STM) based techniques [6,7], mechanically controlled break junction methods (MCBJ) [8], or conductive probe atomic force microscopy (C-AFM) [9]. Large numbers of the resulting traces are then collected together and analyzed to then provide insight into the junction electrical characteristics.

In cases where junction formation probabilities are high, and particularly with "simpler" molecular wires which employ effective contacting groups, conductance features can be generally readily recognized in histograms which include all junction formation cycles. However, this is not always the case, as junction formation probabilities (JFPs) can be low and there can be a large dispersion in junction configurations, particularly for more complex systems. Junction formation probabilities can also be method dependent. For example, the STM-BJ in which the STM tip is crashed into the substrate at the start of every junction formation cycle can show much higher JFPs in a range of 70-95% [10] than "non-contact" methods, such as the $I(s)$

method which avoids tip-to-substrate contact. The differing JFPs of the STM-BJ and $I(s)$ methods arise from different ways in which junction formation is achieved. The STM-BJ method forms and cleaves an Au-Au contact before forming the Au-molecule-Au junction. On the other hand, in the $I(s)$ technique the STM tip is approached to the substrate at constant bias voltage to reach a pre-determined set point tunnelling current I_0, which is set to below the current which would result in tip-to-substrate contact. The STM tip is then pulled away from the substrate, while the tunnelling current (I) is recorded as a function of the tip/substrate distance (s). The accumulated I-s traces then provide a gateway to determining the molecular conductance (G), the junction break-off distance and also the decay constant (β) if a homologous series of molecules is measured [7,11–12].

The typically low JFPs of the STM-$I(s)$ technique have meant that this technique has typically relied on data sorting or selection procedures, particularly to remove the current-distance traces which are "blank" traces in which there is no junction formation. Such traces display an exponential current decay or are excessively noisy, for example as a result of poorly contacted molecular wires. Early on in the technique's innovation such "filtering" was achieved through laborious hand-selection, although it was recognized that this could introduce user bias and limit the size of the datasets [13]. This conventional data analysis relied on manual data selection through the recognition of traces containing plateaus. The selected current-distance traces were then used to construct conductance histograms, with the most probable conductance value(s) being obtained from the position of the conductance peak [14]. Such procedures require prior signal shape identification, and it is therefore difficult to completely avoid assumptions about the expected outcome when the signal shape presents more variability. In addition, some data features may remain unnoticed leading to loss of information [15].

In an attempt to ameliorate the influence of these issues in data-analysis, unsupervised data-analysis algorithms have been proposed, which involve computer processing and sorting of the current-distance or current-voltage data. These have used a number of automatic data analysis methods and algorithms to filter, sort and display

the electrical data for large molecular junction datasets [5,15–21]. An early strategy to achieve this was proposed by Jang *et al*. [22], in a method that they called a "last-step analysis" (LSA). They applied this analysis to the conductance behavior of series of thiol-terminated alkanes with the STM-BJ technique. In their last step analysis method, only the final rapid drop of a conductance trace was used to plot the conductance histogram. This procedure does not require any manual data preselection making the results less subjective. As discussed in the preceding text compared to the STM-BJ, data processing for STM-*I*(s) technique can be even more difficult because of the lower junction formation probabilities. Here, Albrecht's group have reported a specific algorithm for the unsupervised data processing of large *I*(s) datasets [13]. In this algorithm, each *I*(s) trace is divided into many bins with a selected bin width (BW). Each bin will have a value for the number of plateaus-determining bin counts. Using a sorting algorithm that recognizes plateau containing traces from the bin counts, the traces containing plateau (molecular) features can be separated from both exponential traces and those primarily related to the experimental noise. Their sorting algorithm and program offered a robust method to identify the appropriate plateau feature contained traces associated with the most probable junction configuration

Machine learning and artificial intelligence techniques have also been applied to recognize and classify features within single-molecule break junction datasets [16,21,23–26]. Machine learning based approaches are well suited for the analysis of large experimental datasets without prior knowledge and can help to significantly reduce the bias [27]. For instance, Lauritzen et al. [26] reported a data analysis method that using recurrent neural networks (RNNs) trained on the traces of minimal cross-sectional area, which could be deemed as a proxy for conductance in molecular dynamics simulations. Based on this powerful technique, they separate two sets of conductance traces. The machine learning approaches are capable of representing and analyzing multiple aspects of measured data, but are ideally used in conjunction with large amounts of data and training.

In 2018, our group also reported an efficient data sorting algorithm which was

specifically developed for the STM-*I*(s) measurement data with minimal training and computing overhead [17]. The automatic data analysis process focuses on the removal of traces where proper molecule junction formation does not occur. The remaining traces pertaining to effective junction formation are then displayed using a conductance mapping process which makes it possible to analyse the statistical diversity in the dataset and recognize groups and sub-groups in the data. This captures the statistical complexity of the molecular system in a straightforward way, enabling analysis of the effects of molecular structure, contacts, environment and other physical parameters. This algorithm has been validated through measurements of the model Au/1,8-octanedithiol/Au system [12]. In this work, we expand the application scope of this data sorting algorithm to four more complex molecular junction systems. The sorting algorithm is evaluated against 4 different molecules, presenting differences in conjugation (saturated versus π-conjugated molecular bridges), contacting groups (thioanisole, thiomethyl and isothiocyanate used), length (biphenyl and terphenyl bridges used) and targets presenting different anchoring groups at either terminus. This diversity of molecular junctions helps us to ascertain the effectiveness of the sorting algorithm across a diverse range of molecular targets. Our experimental results show that this algorithm deals well with the variability of conductance dispersions across large datasets and offers new insights into the studied molecular junctions.

Experimental

In this research, four different molecules were investigated in gold-gold electrode junctions, which are chain-like 6-(methylthio) hexyl isothiocyanate (6MHI) and 8-(methylthio) octyl isothiocyanate (8MOI), and bi- and ter-phenyl compounds terminated symmetrically by thiol groups, including biphenyl-4,4'-dithiol (DBDT) and p-terphenyl-4,4''-dithiol (TBDT). The respective molecular structures are shown in Figure 1a. In the measurements, 1 mM solutions of the target molecule in mesitylene (99%) were prepared for each case. All these used molecules are commercially obtained. Conductance measurements were performed using the previously reported STM-*I*(s)

method [7] in ambient conditions. The conductance measurements were performed with 300 mV bias voltage for 6MHI and 8MOI molecular systems, and 100 mV bias voltage for DBDT and TBDT molecular systems. The schematic illustration of a single molecular junction with Au/Au contacts is shown in Figure 1b. All analytes and solvents were obtained from commercial sources.

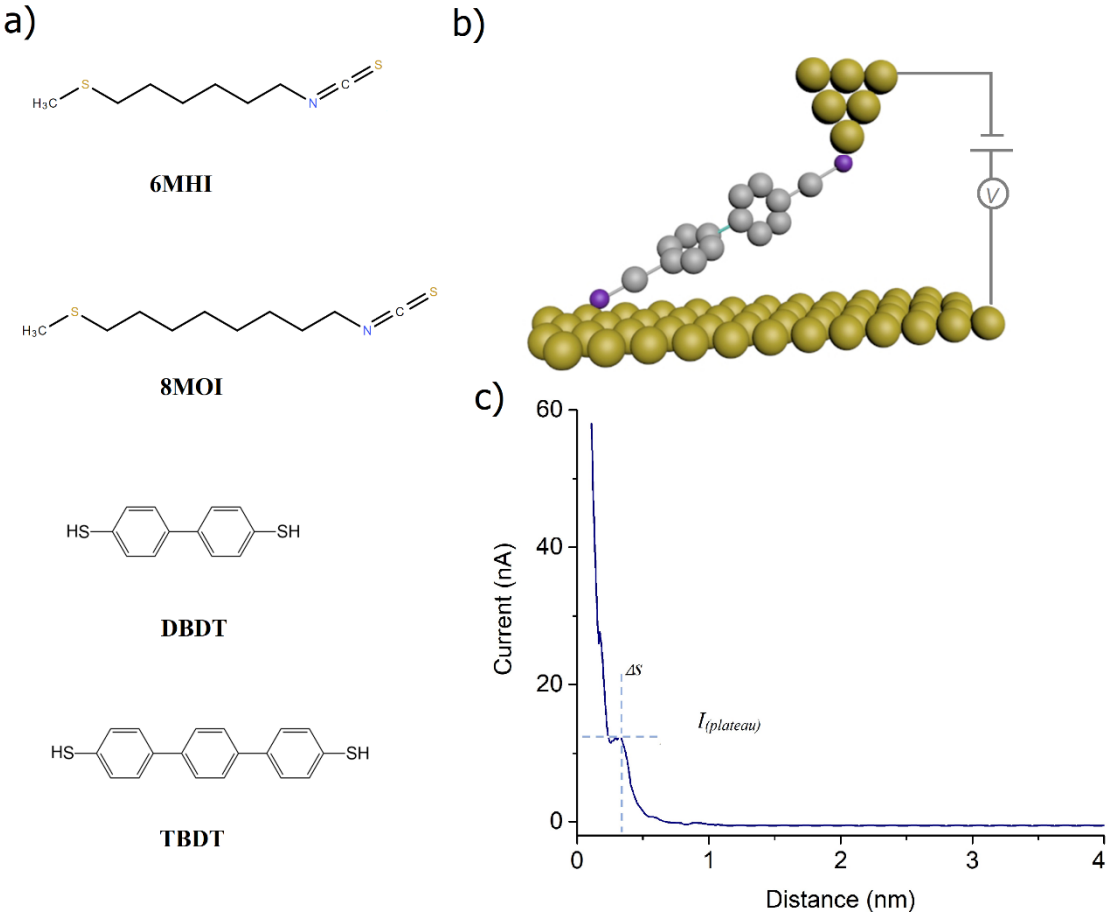

Figure 1: (a) Molecular formula of measured molecules: 6-(methylthio) hexyl isothiocyanate (6MHI), 8-(methylthio) octyl isothiocyanate (8MOI), biphenyl-4,4'-dithiol (DBDT) and p-terphenyl-4,4''-dithiol (TBDT); (b) A schematic illustration of a single molecular junction formed by STM-*I*(*s*) technique; (c) A characteristic *I*(s) trace showing a plateau for Au/DBDT/Au.

All the data processing used the MATLAB algorithm reported by our group in 2018. [17] The *I*(s) traces recorded with the STM instrumentation can be exported into a single ASCII document, which contains distance and current signals in *n* rows, as illustrated in Figure 2a. For each target molecule, over 30000 *I*(s) traces were measured

and collected. This algorithm firstly removes noisy *I*(s) traces. Figure 2b shows typical *I*(s) traces. The green lines feature plateau containing traces, which is a sign of the formation of a molecular junction. In such traces, following the plateau the tunneling current decays to zero with increasing tip-substrate distance as the molecular bridge is broken.[28] The blue lines show a simple current decay with no molecular junction being formed. The red lines in Figure 2b represent three types of noisy traces, which can be filtered out by the X-filter, Y-filter, and peak-filter algorithm respectively, as previously described. The X-filter algorithm removes current traces which does not smoothly decay to zero during junction cleavage or where noise features appear in the low current range. The Y-filter algorithm removes *I*(s) traces which exhibit an oscillating or excessively noisy signal at the beginning of the current decay, which may arise from poorly contacted molecules, contaminated tips or multiple interacting molecules in the junction. The peak-filter algorithm helps to remove the undesired noisy traces, for example with excessive instability and associated noise in the plateau region, which would otherwise contribute to a deleteriously broad peak in the 1D conductance histogram.

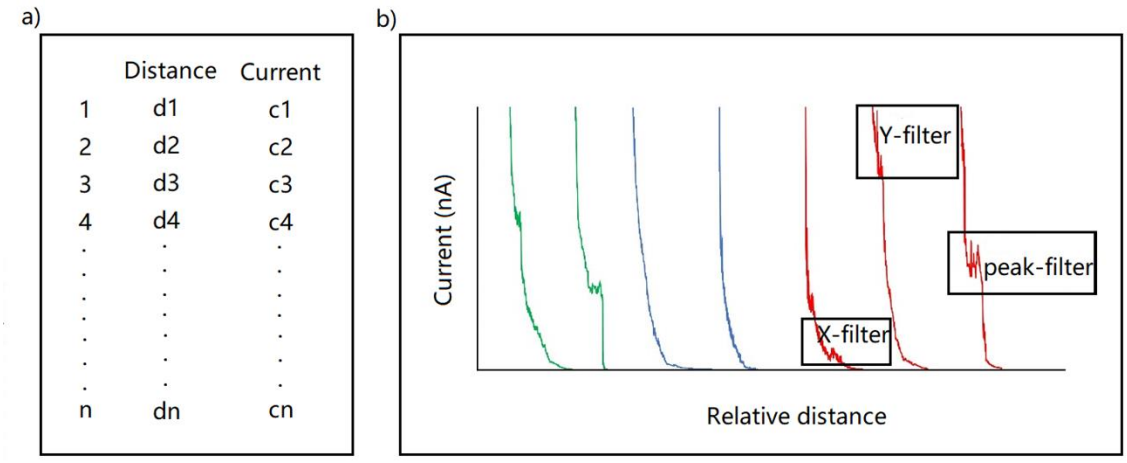

Figure 2: (a) Data structure of *I*(s) data; (b) Green: plateau featuring *I*(s) traces; blue: decay traces without molecular junction formation, red: noisy traces which would be removed by X-filter, Y-filter and Peak-filter algorithms, respectively.

A conductance mapping algorithm was then applied to the experimental data. The conductance mapping process is used to obtain the most dominant conductance values for a given molecular target. To illustrate how this conductance mapping operates data from Au/ 6MHI /Au molecular junction formation is used as an example. For each single *I*(s) trace the Y-axis (current axis) is simply divided into many steps (bins). As we discussed in the previous work, the bin numbers do not rely on the subjective choices of the operator, but are based on a trial and error scheme [17]. The counts of the data points are different in each bin, and the plateau region will always have a larger bin count than adjacent bins that do not lie in the plateau region. Thus, the maximum number of bin counts can be regarded as a criterion to separate the pure decay curve from the plateau featuring curves. During the selection process, only two parameters need to be set, one is the bin width (how many steps to divide the Y axis of current signal), while the other one is the bin count number (how many data points fall in this bin width). An appropriate setting is achieved after obtaining optimised 1D and 2D histograms. Notably, the software can list and try many possible bin width and bin count combinations in a very short time. For further explanations of the algorithm see reference [17]. As illustrated in Figure 3a, it seems that the plateaus are distributed in a few different groups. The conductance map is then divided into three regions based on the plateau counts and these are color coded red, green and blue in Figure 3b.

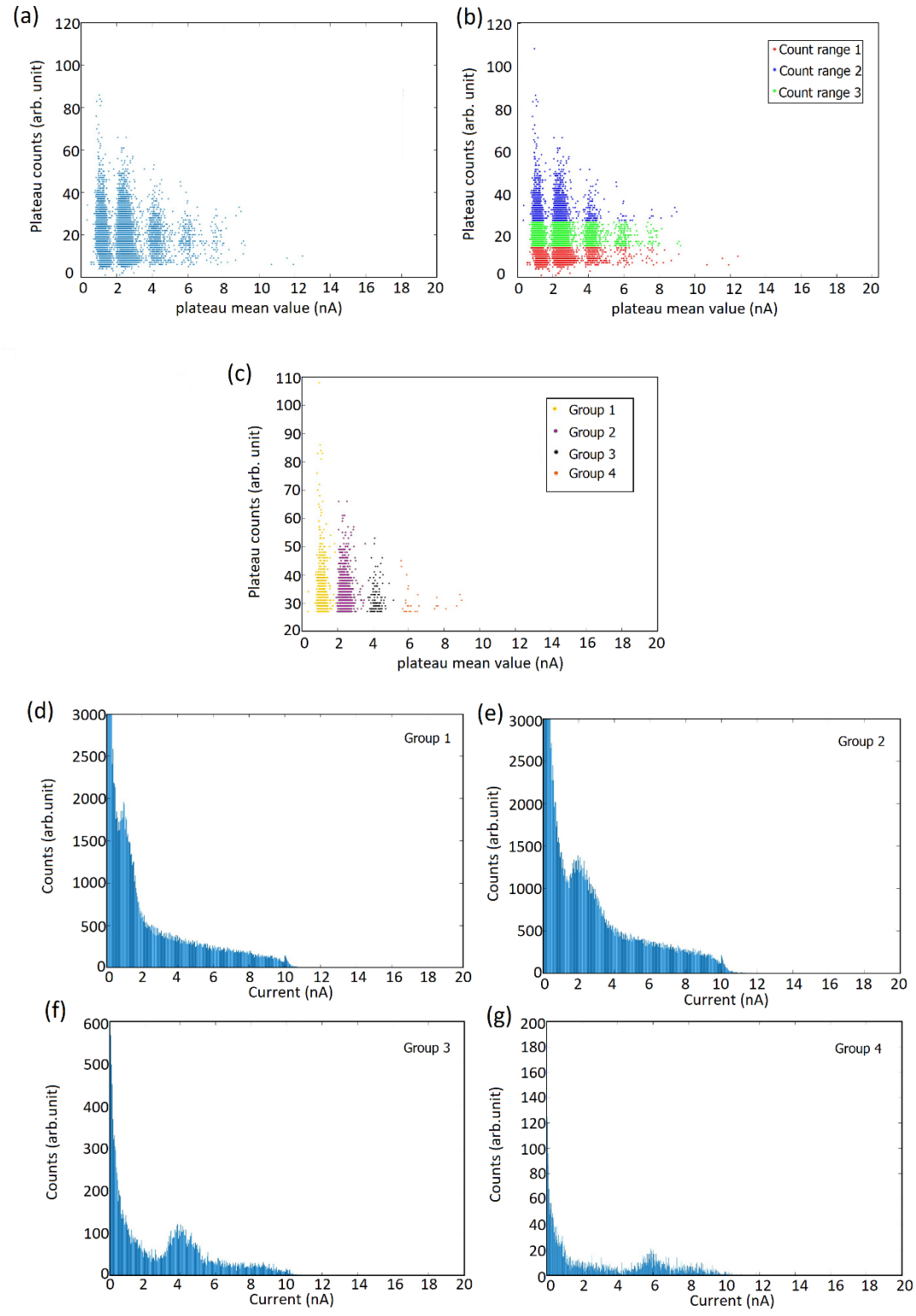

Figure 3: Data automatically generated by the algorithm for Au|6MHI|Au molecular junction formation. (a) Current mapping grouped for Au|6MHI|Au in five plateau mean value regions to get insights in the most dominant peak; (b) Current mapping for Au|6MHI|Au with color-coded grouping into three count ranges based on the plateau counts; (c) Current mapping for Au| 6MHI|Au of the ideal plateau region , four refined groups which are color coded represent the most dominant conductance peak and sub-

groups; (d) 1D current histogram plotted with *I*(s) traces from the group 1 (yellow in Figure 3c; (e) 1D current histogram plotted by *I*(s) traces from the group 2 (purple) in Figure 3c; (f) 1D current histogram plotted with *I*(s) traces from the group 3 (black) in Figure 3c; (g) 1D current histogram plotted with *I*(s) traces from the group 4 (orange) in Figure 3c. After determining the dominant current groups in (d)-(g), the peak conductance can be determined from $G = \frac{I}{V_{bias}}$.

As shown in Figure 3b, count range 1 (red) represents very small count values as occurs, for example in the steep decay region of the *I*-s curve with no plateau features. Count range 2 in Figure 3b (green) reflects the *I*(s) curves with short plateau or small noisy peaks, while group 3 (blue) is related to the plateau region with plateau curves of sufficient extension. Based on the conductance mapping function, optimized plateau distribution mapping was obtained next. As shown in Figure 3c, for Au|6MHI|Au molecular junction, there are four separable plateau distribution regions, which are distinguished by different colors. These are referred to as different "conductance groups".

This algorithm then plotted a 1D current histogram for each conductance region (group) automatically, as shown in Figure 3d-g. Figure 3d corresponds to group 1 (yellow) in Figure 3c. Figure 3e corresponds to group 2 (purple) in Figure 3c. Figure 3f corresponds to group 3 (black) in Figure 3c, and Figure 3e corresponds to group 4 (orange) in Figure 3c. The region (group) having the most data points is considered as the most probable (dominant) conductance group, while the other regions would be considered as the sub-group. In this measurement, group 1 contains 1021 *I*(s) curves, group 2 contains 1401 *I*(s) curves, group 3 contains 126 *I*(s) curves, and group 4 contains 28 *I*(s) curves. The Au/6MHI/Au molecular junction group 2 (purple) in Figure 3c was the dominant conductance group.

To analyze further the molecular junctions, an estimated total break-off distance S_{total} can be calculated using equation 1. S_{total} is usually compared to the length of the fully extended molecular junction, typically obtained by molecular modelling. If a comparable value is obtained between the theoretical molecular length and

experimental S_{total} value, this indicates a molecule successfully bridging the gap between the two gold electrodes.

$$S_{total} = S_0 + \Delta S \quad (1)$$

In this equation, ΔS represents the distance retracted by the STM tip from the set point current distance to the junction break-off point, after which the plateau decays smoothly to zero current as shown in Figure 1c (where the two dotted lines intersect). The experimental ΔS can be calculated for each molecular system using a statistical analysis. While S_0 is the distance between the STM tip and substrate at a predetermined set point current (I_0).[30] S_0 can be calculated by the following equation:[30]

$$S_0 = \frac{\ln(G_0 \times \frac{V_{bias}}{I_0})}{\frac{d\ln(I)}{ds}} \quad (2)$$

Where $G_0 = 77400\ nS$, V_{bias} is the applied voltage between tip and substrate. $\frac{d\ln(I)}{ds}$ can be obtained as the slope of ln(I) versus distance (I(s) traces are selected from those where no molecule is linked between electrodes). For each of molecular system, we calculated 30 $\frac{d\ln(I)}{ds}$ values to further obtain their average value. Here, the S_0 data were recorded as 0.38 nm for Au|DBDT|Au, 0.41 nm for Au|TDBD|Au, 0.34nm for Au|6MHI|Au isothiocyanate, and 0.36 nm for Au|8MOI|Au molecular junctions.

Computational

In order to better characterize the molecular junctions considered in this work, we have performed electronic structure calculations using Density Functional Theory (DFT) and electronic transport calculations using a non-equilibrium Green's function (NEGF) formalism within the Fisher-Lee approach.[14] DFT calculations have been performed using the very efficient localized-orbital basis set code Fireball.[14] A basis sets of sp^3d^5 numerical orbitals for Au, sp^3 for N, C and S, and s for H have been used to determine the atomic configuration and the electronic properties of the junctions. The corresponding cut-off radii (in atomic units) are s = 4.6, p = 5.2, d = 4.1 (Au), s = 4.2, p = 4.2 (N), s = 4.5, p = 4.5 (C), s = 3.1, p = 3.9 (S), and s =4.1 (H). Following a previously developed procedure,[12] we have used gold electrodes of 35 atoms. This

code uses the local density approximation (LDA) through the McWeda formalism[36]. All the junctions have been optimized using DFT until the forces went below 0.05 eV/Å.

Results and discussion

For Au|6MHI|Au molecular system, we replotted the 1D conductance histogram with the *I*(s) traces from the dominant region as shown in Figure 4a, where the peak position indicates a conductance value is 6.6 nS. Figure 4b shows the distribution of ΔS values for all the measurements, the average experimental break off value is found to be $0.6 \pm 0.3nm$, where the error bar was calculated as the full width at half maximum (FWHM) of the peak fitting in Figure 4b, the relative S_{total} could be obtained as $0.94 \pm$ 0.3 nm which agrees well with the molecular size.

To provide further insight into the experimental results, a conductance histogram has been plotted for group 1 of Au|6MHI|Au, which is the region containing the second highest plateau data count, as shown in Figure 4c. The conductance value from this group 1 of Au|6MHI|Au molecular junctions is 3.2 nS. The ideal molecular length for 6MHI (from sulfur atom to sulfur atom) is 1.12 nm, the relative break-off distance is found to be 0.65 nm, which gives an S_{total} value of 0.99 nm. This result agrees well to the molecular junction length of 1.01 nm (estimated using ChemBio3D Ultra 20.0) when the isothiocyanate group tilts on the gold electrode and the relative molecular junction length was calculated from one side gold electrode atom to another side N atom. We notice that similar S_{total} is recorded for group 1 and group 2. This suggests that the different conductance value recorded between these two different groups might result from different surface binding geometries with similar overall junction lengths.

We subsequently plotted the conductance histogram obtained from group 3. There are only 126 valid *I*(s) curves contained in that region. Compared to group 1 and 2, the data volume of group 3 is dramatically reduced. The 1D conductance histogram of group 3 is shown in Figure 4e, with the peak position giving the higher conductance value of 13.3 nS. The break-off distance calculation reveals more details of the molecular junctions accounting for group 3. As shown in Figure 4f, the average

experimental break off value obtained from group 3 is 0.39 nm, giving an S_{total} of 0.73 nm. Importantly, this S_{total} value is considerably smaller than the 6MHI molecular length. If the 6MHI molecular bridge is tilted between the gold contacts at the point of junction cleavage then the break-off distance would be decreased. In addition, in this kind of configuration electrons could tunnel through the molecular junction without having to tunnel through the whole molecular backbone, thereby “short-circuiting" through bond transport which could effectively reduce the junction resistance. These experimental findings support the fact that higher conductance of the molecular junction could be driven by the tilted bridging molecules in the contact gap.

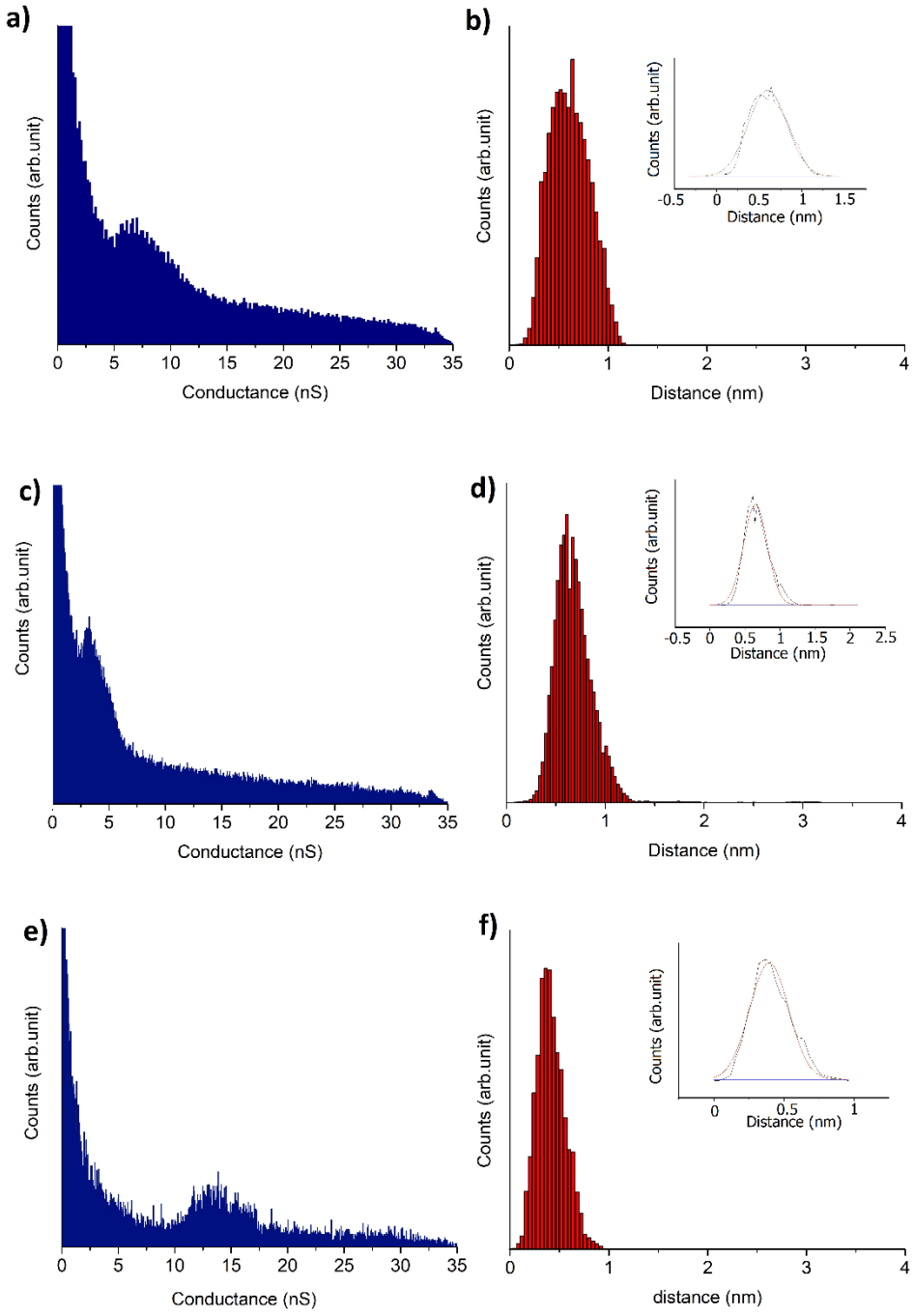

Figure 4: (a) 1D conductance histogram of Au|6MHI|Au (group 2, the dominant region); (b) Break off distance histogram for the molecular junctions, where the inset is the gaussian peak fitting; (c) 1D conductance histogram of Au|6MHI| Au (group 1); (d) Break off distance histograms; (e) 1D conductance histogram of Au|6MHI|Au (group 3); (f) Break off distance histograms.

The mapping results of Au|8MOI|Au molecular junctions, automatically generated by the algorithm, are shown in Figure 5a and b. There are four distinct conductance groups

as shown in Figure 5b, where region 1 (red) is the dominant one. The 1D conductance histogram is shown in Figure 5c, where the peak position lies at a conductance value of 2.2 nS. Figure 5d shows the distribution of ΔS for all the measurements, the average experimental break off value is $0.33 \pm 0.2nm$, and the relative S_{total} is $0.74 \pm 0.2nm$. The measured break-off distance is smaller than the theoretical molecular size of 8MOI. This result indicates an unexpected tilt of the bridging molecules between the STM tip and substrate at the average junction breaking length. Such a tilting of the bridging molecule could lead to a higher measured conductance value.

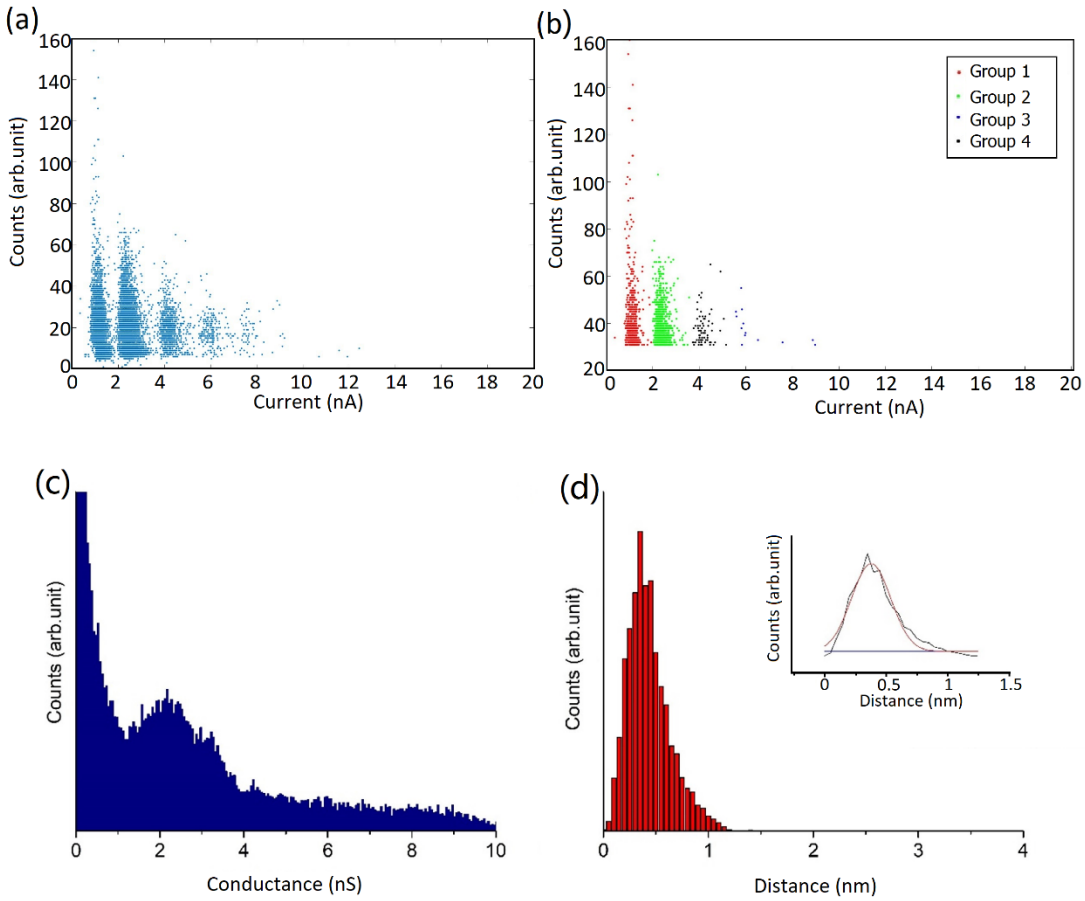

Figure 5: Data automatically generated by algorithm (a) Conductance mapping for of Au|8MOI |Au molecular junctions grouped into few regions after removing the noisy *I*(s) traces. (b) Conductance mapping of the ideal plateau region, three refined groups were identified as the most dominant conductance peak and sub-groups, where group (1) red is the most possible conductance region. (c) The 1D conductance histogram plotted by the *I*(s) traces that came from group 1 in Figure 5b; (d) Break off distance histogram, where the inset is the gaussian peak fitting.

We further compared this asymmetric $\mathrm{Au|CH_3S-(CH_2)_n-NCS|Au}$ molecular junction to the previous reported symmetric $\mathrm{Au|SCN-(CH_2)_n-NCS|Au}$ junction [31]. Conductance values have been reported as 15.5 nS ($2\times10^{-4}G_0$) for $\mathrm{Au|NCS-(CH_2)_6-NCS|Au}$, and 2.6 nS ($0.34\times10^{-4}G_0$) for $\mathrm{Au|SCN-(CH_2)_8-NCS|Au}$. This comparison shows that the asymmetric anchoring groups of the $\mathrm{Au|CH_3S-(CH_2)_n-NCS|Au}$ molecular junction lead to a lower conductance than for the symmetric analogue, $\mathrm{Au|SCN-(CH_2)_n-NCS|Au}$. It is worth noting that it has been previously found for asymmetric anchoring group systems, that while the conductance decreases with the molecular length, the attenuation with length is slower than the $\mathrm{Au|NCS-(CH_2)_n-NCS|Au}$ [31]. This conductivity decreases due to the asymmetry of the molecular junction had been reported in our previous research [12] which has been rationalized by the breaking of junction symmetry, the variation of electrode|molecule coupling and the energy level alignment.

The mapping results of Au|DBDT|Au and Au|TBDT|Au junctions automatically generated by the algorithm are shown in Figure 6. For the Au|DBDT|Au junction, there are four main distinct conductance groups, where the group 1 (in Figure 6b) represents the dominant conductance group. The rough conductance behavior of Au|TBDT|Au junctions is shown in Figure 6c. There are four refined conductance groups in Figure 6d, where group 2 is the dominant conductance group.

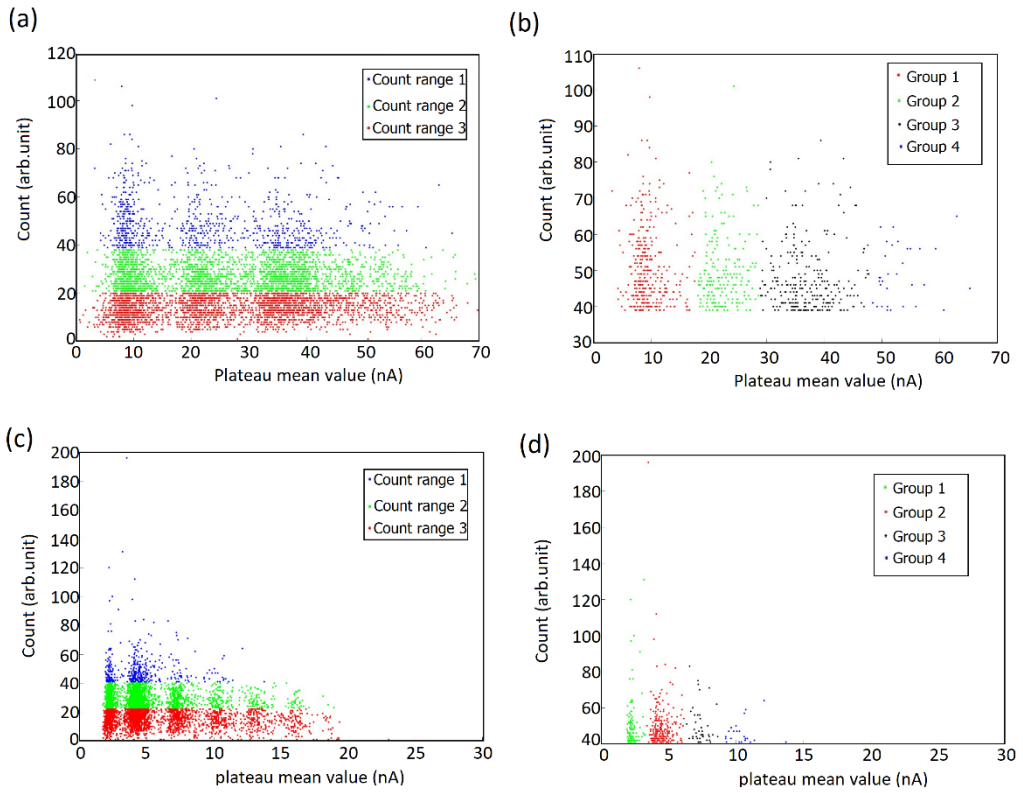

Figure 6: Data automatically generated by the algorithm: (a) Conductance mapping for Au|DBDT|Au molecular junctions with color-coded grouping into three main regions based on the plateau counts. (b) Conductance mapping of the ideal plateau region, four refined regions were grouped to examine the most dominant conductance peak and sub-groups, where group 1 is the dominant conductance group. (c) Conductance mapping for Au|TBDT|Au molecular junctions with color-coded grouping into three main regions based on the plateau counts. (d) Conductance mapping of the ideal plateau regions, four refined regions were grouped to examine the most dominant conductance peak and sub-groups, where group 2 is the dominant group.

The dominant 1D conductance histogram of Au|DBDT|Au, is shown in Figure 7a, and this is plotted from the *I*(s) traces from group 2 in Figure 6b, where the conductance is recorded as 118 nS following a gaussian peak fitting. Figure 7b shows the distribution of ΔS for all the measurements, the peak position is $0.87 \pm 0.4nm$, so the relative S_{total} is obtained as $1.25 \pm 0.4nm$, which agrees well with an ideal molecular junction length of 1.29 nm (estimated using ChemBio3D Ultra 20.0). This result proves the bridged DBDT molecule is successfully linked between two gold electrodes, while the Au-S bonds are formed on each side.

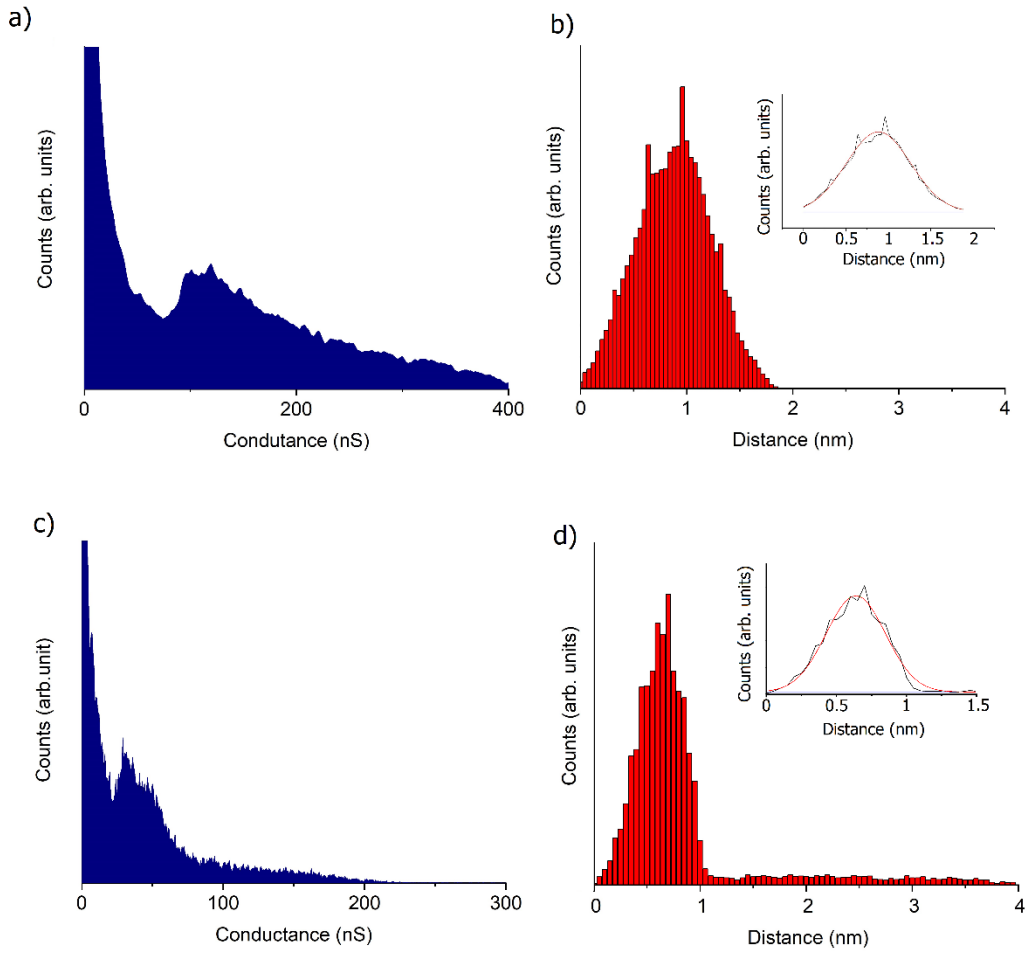

Figure 7: (a) 1D conductance histogram of Au|DBDT|Au; (b) Break-off distance histograms for the molecular junctions, where the inset is the Gauss peak fitting; (c) 1D conductance histogram of Au|TBDT|Au; (d) Break-off distance histograms, where the inset is the Gauss peak fitting.

The dominant 1D conductance histogram of Au|TBDT|Au, as shown in Figure 7c, which is plotted from the *I*(s) traces from group 2 in Figure 6d, where the conductance was recorded as 32 nS after a gaussian peak fitting. Figure 7d shows the distribution of ΔS from this measurement, where the peak position is $0.64 \pm 0.25\ nm$, so the relative S_{total} could be obtained as $1.05 \pm 0.3\ nm$. This measured break-off distance is smaller than the molecular size of TBDT, and it is smaller than the break-off distance of the Au|DBDT|Au system. This result indicates an unexpected tilt of TDBT bridging between the STM tip and substrate. In addition, we found that, the conductance values

(118 nS and 32 nS) of Au|DBDT|Au and Au|TDBT|Au junctions obtained using the MATLAB algorithm agree within acceptable limits with the manual selected conductance values (126 nS and 23 nS) we reported in previous research [14] suggesting reliable attributes of this data selection algorithm.

The specific conductance values of Au/6MHI/Au, Au/8MOI/Au, Au/DBDT/Au and Au/TBDT/Au are listed in Table 1. By comparing the electrical properties of molecular bridges containing saturated polymethylene chains with the oligo-phenyl bridged molecular junctions, it is clear that the conductances of oligo-phenyl based molecular junctions are significantly greater than polymethylene backbone molecular junctions even when they have similar molecular length. This finding is completely in line with the expectations that the conjugated backbones have greatly enhanced electrical conductance compared to the non-conjugated bridges.

Table 1: The conductance values and break-off distances of Au/6MHI/Au, Au/8MOI/Au, Au/DBDT/Au and Au/TBDT/Au systems based on unsupervised data analysis.

Molecule	**Ideal molecular junction length**	**Dominant conductance (nS)**	S_w**(nm)**	**Experimental break-off distance (nm)**
6MHI	1.12	6.6	0.60 ± 0.3	0.94 ± 0.3
8MOI	1.36	2.2	0.38 ± 0.2	0.74 ± 0.2
DBDT	1.29	118	0.87 ± 0.4	1.25 ± 0.4
TBDT	1.69	32	0.64 ± 0.3	1.05 ± 0.3

In order to analyze the influence of a variety of plausible atomic configurations in

the studied molecular junctions, we have performed electronic structure calculations using Density Functional Theory (DFT) and electronic transport calculations using a non-equilibrium Green's function (NEGF) formalism within the Fisher-Lee approach [32,33]. DFT calculations have been performed using the very efficient localized-orbital basis set code Fireball [34,35]. This code uses the local density approximation (LDA) through the McWeda formalism [36]. First, we have considered different possible TBDT molecular junction configurations, including molecular junctions with two TBDT molecules. The corresponding atomic configurations are represented in Figure 8a for the representative configurations used in the modelling.

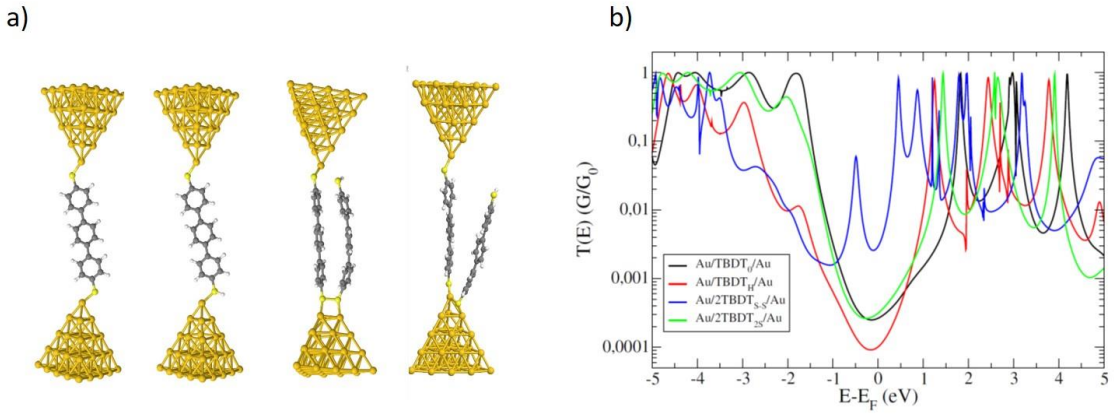

Figure 8: (a) DFT-optimized possible atomic configurations of several TBDT-based molecular junctions. From left to right, (i) standard TBDT junction with both thiol groups deprotonated (i.e. Au-S bonds), (ii) TBDT junction with one connection through a deprotonated thiol and the other connection retaining the H atom on the thiol, (iii) one TBDT molecule with a neighboring one connected through a sulfur bridge and finally (iv) one TBDT molecule with a neighboring TBDT connected to the same electrode. (b) Corresponding electronic transmissions.

From left to right, we have (i) a standard TBDT molecular junction with connection of the TBDT to the gold electrodes through sulfur atoms forming Au-S chemisorption bonds, (ii) the same junction but with one thiol retaining its hydrogen atom while the other is deprotonated, (iii) one molecule connected to the gold electrode with a neighboring molecule connected to the sulfur atom forming a disulfide bridge, and finally (iv) one molecule connected but with a neighboring molecule connected

alongside it as illustrated. The sulfur-sulfur binding configuration (disulfide bond between neighboring molecules) was proposed in a previous study, which suggested that disulfide-mediated dimerization of BDT contributes to an experimentally observed low conductance feature.[37] In the present work, the main objective of these calculations is to show that several different configurations may lead to similar conductance measurements. In addition, complex configurations are likely to occur experimentally when an STM tip penetrates a dense TBDT self-assembled monolayer, since opportunities exist to pick up multiple molecules from such dense molecule arrays. After having optimizing the geometry of these configurations, the corresponding electronic transmissions were calculated. The results are presented in Figure 8b in electronic transmission plots of $T(E)$ versus E-E_F. As can be observed, although the transmission curves are all highly different it is still possible to get very similar transmission values, at the Fermi level, for very different junction configurations. Namely, the black and green transmission practically coincide around the Fermi energy. The black curve corresponds to the standard deprotonated thiol configuration while the green curve corresponds to the arrangement with a neighboring molecule connected to one electrode besides the main molecular bridge. These two strikingly different configurations present roughly the same conductance, of 20.6 and 24.6 nS, respectively. We think there are two possible explanations for this similarity in the conductance: first, the overall configuration is a bit different and the atomic positions are slightly different due to the close presence of the non-bridging TBDT. Second, due to the proximity of this second molecule, the electrostatics of the first molecule is slightly modified, as the result of a close molecular dipole for example. In any case, we only presented here a possible configuration, which could lead to different conductance plateaus, as observed experimentally, but this configuration might not be realistic. On another hand, the junction involving two molecules and a sulfur bridge presents a very high conductance, 234 nS, while the junction involving the connection through a thiol group presents a very low conductance of 7.9 nS. These results give further insight into the experimental configuration by affirming the different junction configuration can produce markedly

different conductance groups, which justifies the multiple groups observed after the data sorting algorithm, but it also shows that care is needed since it is also possible for different configurations to having coinciding conductance. These findings show the importance that both a data sorting analysis complemented with theoretical computations are needed to gain insight into the influence of different configurations on junction electrical properties.

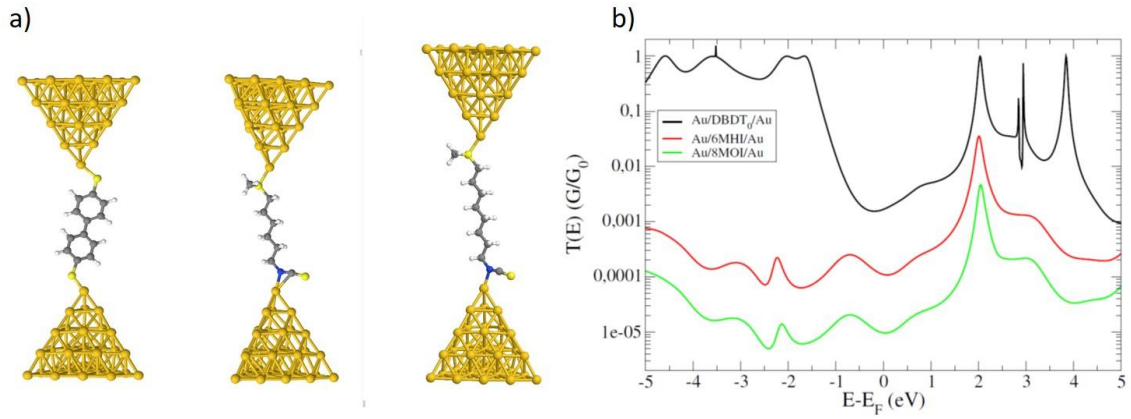

Figure 9: (a) From left to right, DFT-optimized atomic configurations of DBDT, 8MOI and 6MHI molecular junctions; (b) Corresponding electronic transmissions.

In the same manner, we have also analyzed DBDT, 6MHI and 8MOI molecular junctions. The corresponding DFT-optimized atomic configurations are represented in Figure 9a. As already observed experimentally, the 6MHI and 8MOI containing junctions exhibit much lower conductance than their TBDT counterpart despite similar lengths, with conductance value of 8.5 and 0.75 nS, respectively. By comparing the conductance behavior of these two groups of molecular junction systems, we found that increased extent of conjugation can lead to much higher conductance. As expected, the DBDT molecular junction presents a higher conductance, of 130 nS, than the TDBT junctions, due to the longer molecular length of the latter.

Conclusion

In this study, the conductance of four molecular junctions with Au|Au contacts has

been investigated using STM-*I*(s) technique. A fully automated MATLAB data sorting algorithm was used to analyze the most dominant conductance groups for Au|6MHI|Au, Au|8MOI|Au, Au|DBDT|Au and Au|TBDT|Au systems. Summarizing the experimental results detailed above, the following general observations can be made. For each molecular system studied the sorting algorithm is able to distinguish multiple conductance groups and highlight the dominant one. The data sorting algorithm provides conductance values which are consistent with literature and expectations, with the dominant conductance group of the conjugated molecular bridges showing significantly higher conductance than the similar length non-conjugated molecular bridges. Analysis of the junction breaking distances of separate groups can give insights into possible conformational differences between different conductance groups, as shown in particular for the example of Au|6MHI|Au junctions. It is additionally noted that the measured break-off distance values of Au|8MOI|Au and Au|TBDT|Au are shorter than their respective molecular length, which indicates a possible tilted configuration of the bridging molecules at the point of junction breaking.

Conflict of interest

The authors declare no competing financial interest.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC Grants 21503169), Suzhou Industrial Park Initiative Platform Development for Suzhou Municipal Key Lab for New Energy Technology (RR0140), Key Program Special Fund in XJTLU (KSF-E-38) and the XJTLU Research Development Fund (RDF-16-01-33 and REF-19-01-05).

References

[1] L. Lafferentz, F. Ample, H. Yu, S. Hecht, C. Joachim, L. Grill, Conductance of a Single Conjugated Polymer as a Continuous Function of Its Length, Science. 323 (2009) 1193–1197.

[2] N. J. Tao, Electron transport in molecular junctions, Nature Nanotechnology. 1 (2006) 173–181.

[3] E. A. Weiss, M. R. Wasielewski, M. A. Ratner, Molecules as Wires: Molecule-Assisted Movement of Charge and Energy, in: Mol. Wires Electron., Springer, Berlin, Heidelberg, 2005: pp. 103–133.
[4] X. Y. Xiao, B. Q. Xu, N. J. Tao, Conductance Titration of Single-Peptide Molecules, J. Am. Chem. Soc. 126 (2004) 5370–5371.
[5] A. Vladyka, T. Albrecht, Unsupervised classification of single-molecule data with autoencoders and transfer learning, Mach. Learn. Sci. Technol. 1 (2020) 035013.
[6] B. Xu, N. Tao, Measurement of Single-Molecule Resistance by Repeated Formation of Molecular Junctions, Science. 301 (2003) 1221–3.
[7] W. Haiss, H. van Zalinge, S. J. Higgins, D. Bethell, H. Höbenreich, D. J. Schiffrin, R. J. Nichols, Redox state dependence of single molecule conductivity, J. Am. Chem. Soc. 125 (2003) 15294–15295.
[8] M. A. Reed, C. Zhou, C. J. Muller, T. P. Burgin, J. M. Tour, Conductance of a Molecular Junction, Science. 278 (1997) 252–254.
[9] X. D. Cui, A. Primak, X. Zarate, J. Tomfohr, O. F. Sankey, A. L. Moore, T. A. Moore, D. Gust, G. Harris, S. M. Lindsay, Reproducible measurement of single-molecule conductivity, Science. 294 (2001) 571–574.
[10] T. Fu, K. Frommer, C. Nuckolls, L. Venkataraman, Single-Molecule Junction Formation in Break-Junction Measurements, J. Phys. Chem. Lett. 12 (2021), 10802-10807.
[11] R. J. Nichols, W. Haiss, S. J. Higgins, E. Leary, S. Martin, D. Bethell, The experimental determination of the conductance of single molecules, Phys. Chem. Chem. Phys. 12 (2010) 2801–2815.
[12] Q. Zhang, L. Liu, S. Tao, C. Wang, C. Zhao, C. González, Y. J. Dappe, R. J. Nichols, L. Yang, Graphene as a Promising Electrode for Low-Current Attenuation in Nonsymmetric Molecular Junctions, Nano Lett. 16 (2016) 6534–6540.
[13] M. S. Inkpen, M. Lemmer, N. Fitzpatrick, D. C. Milan, R. J. Nichols, N. J. Long, T. Albrecht, New Insights into Single-Molecule Junctions Using a Robust, Unsupervised Approach to Data Collection and Analysis, J. Am. Chem. Soc. 137 (2015) 9971–9981.
[14] S. Tao, Q. Zhang, C. He, X. Lin, R. Xie, C. Zhao, C. Zhao, A. Smogunov, Y. J. Dappe, R. J. Nichols, L. Yang, Graphene-Contacted Single Molecular Junctions with Conjugated Molecular Wires, ACS Appl. Nano Mater. 2 (2019) 12–18.
[15] M. Lemmer, M. S. Inkpen, K. Kornysheva, N. J. Long, T. Albrecht, Unsupervised vector-based classification of single-molecule charge transport data, Nat Commun. 7 (2016) 1–10.
[16] N. D. Bamberger, J. A. Ivie, K. N. Parida, D. V. McGrath, O. L. A. Monti, Unsupervised Segmentation-Based Machine Learning as an Advanced Analysis Tool for Single Molecule Break Junction Data, J. Phys. Chem. C. 124 (2020) 18302–18315.
[17] Q. Zhang, C. Liu, S. Tao, R. Yi, W. Su, C. Zhao, C. Zhao, Y. J. Dappe, R. J. Nichols, L. Yang, Fast and straightforward analysis approach of charge transport data in single molecule junctions, Nanotechnology. 29 (2018) 325701.
[18] M. S. Inkpen, M. Lemmer, N. Fitzpatrick, D. C. Milan, R. J. Nichols, N. J. Long, T. Albrecht, New Insights into Single-Molecule Junctions Using a Robust, Unsupervised Approach to Data Collection and Analysis, J. Am. Chem. Soc. 137 (2015) 9971–9981.
[19] L. Lin, C. Tang, G. Dong, Z. Chen, Z. Pan, J. Liu, Y. Yang, J. Shi, R. Ji, W. Hong, Spectral

Clustering to Analyze the Hidden Events in Single-Molecule Break Junctions, J. Phys. Chem. C. 125 (2021) 3623–3630.
[20] N. D. Bamberger, D. Dyer, K. N. Parida, D. V. McGrath, O. L. A. Monti, Grid-Based Correlation Analysis to Identify Rare Quantum Transport Behaviors, J. Phys. Chem. C. 125 (2021) 18297–18307.
[21] T. Fu, Y. Zang, Q. Zou, C. Nuckolls, L. Venkataraman, Using Deep Learning to Identify Molecular Junction Characteristics, Nano Lett. 20 (2020) 3320–3325.
[22] S. Y. Jang, P. Reddy, A. Majumdar, R. A. Segalman, Interpretation of Stochastic Events in Single Molecule Conductance Measurements, Nano Lett. 6 (2006) 2362–2367.
[23] A. Magyarkuti, N. Balogh, Z. Balogh, L. Venkataraman, A. Halbritter, Unsupervised feature recognition in single-molecule break junction data, Nanoscale. 12 (2020) 8355–8363.
[
[24] R. Korol, D. Segal, Machine Learning Prediction of DNA Charge Transport, J. Phys. Chem. B. 123 (2019) 2801–2811.
[25] T. Albrecht, G. Slabaugh, E. Alonso, S. M. M. R. Al-Arif, Deep learning for single-molecule science, Nanotechnology. 28 (2017) 423001.
[26] K. P. Lauritzen, A. Magyarkuti, Z. Balogh, A. Halbritter, G.C. Solomon, Classification of conductance traces with recurrent neural networks, J. Chem. Phys. 148 (2018) 084111.
[27] J. P. A. Ioannidis, Why Most Published Research Findings Are False, CHANCE. 18 (2005) 40–47.
[28] G. Binnig, H. Rohrer, Scanning tunneling microscopy, IBM J. Res. Dev. 44 (2000) 279–293.
[29] G. Sedghi, K. Sawada, L. J. Esdaile, M. Hoffmann, H. L. Anderson, D. Bethell, W. Haiss, S.J. Higgins, R.J. Nichols, Single Molecule Conductance of Porphyrin Wires with Ultralow Attenuation, J. Am. Chem. Soc. 130 (2008) 8582–8583.
[30] L. M. Ballesteros, S. Martín, G. Pera, P. A. Schauer, N. J. Kay, M. C. López, P. J. Low, R. J. Nichols, P. Cea, Directionally Oriented LB Films of an OPE Derivative: Assembly, Characterization, and Electrical Properties, Langmuir. 27 (2011) 3600–3610.
[31] M. D. Fu, I. W. P. Chen, H. C. Lu, C. T. Kuo, W. H. Tseng, C. Chen, Conductance of Alkanediisothiocyanates: Effect of Headgroup−Electrode Contacts, J. Phys. Chem. C. 111 (2007) 11450–11455.
[32] D. S. Fisher, P. A. Lee, Relation between conductivity and transmission matrix, Phys. Rev. B. 23 (1981) 6851–6854.
[33] S. Pitié, M. Seydou, Y. J. Dappe, P. Martin, F. Maurel, J. C. Lacroix, Insights on asymmetric BTB-based molecular junctions: Effect of electrode coupling, Chem. Phys. Lett. 787 (2022) 139273.
[34] J. P. Lewis, P. Jelínek, J. Ortega, A. A. Demkov, D. G. Trabada, B. Haycock, H. Wang, G. Adams, J. K. Tomfohr, E. Abad, H. Wang, D. A. Drabold, Advances and applications in the FIREBALL ab initio tight-binding molecular-dynamics formalism, Phys. Status Solidi B. 248 (2011) 1989–2007.
[35] M. A. Basanta, Y. J. Dappe, P. Jelínek, J. Ortega, Optimized atomic-like orbitals for first-principles tight-binding molecular dynamics, Comput. Mater. Sci. 39 (2007) 759–766.

[36] P. Jelínek, H. Wang, J. P. Lewis, O.F. Sankey, J. Ortega, Multicenter approach to the exchange-correlation interactions in ab initio tight-binding methods, Phys. Rev. B. 71 (2005) 235101.

[37] J. Zheng, J. Liu, Y. Zhuo, R. Li, X. Jin, Y. Yang, Z. B. Chen, J. Shi, Z. Xiao, W. Hong, Z. Tian, Electrical and SERS detection of disulfide-mediated dimerization in single-molecule benzene-1,4-dithiol junctions, Chem. Sci. 9 (2018) 5033–5038.