Unsupervised Segmentation-Based Machine Learning as an Advanced Analysis Tool for Single Molecule Break Junction Data

*Nathan D. Bamberger[†], Jeffrey A. Ivie[†,‡], Keshaba N. Parida[†], Dominic V. McGrath[†], and Oliver L.A. Monti[†,§,]**

[†]Department of Chemistry and Biochemistry, University of Arizona, 1306 E. University Blvd., Tucson, Arizona 85721, USA

[§]Department of Physics, University of Arizona, 1118 E. Fourth Street, Tucson, Arizona 85721, USA

Table of Contents

S.1 MCBJ Set-Up

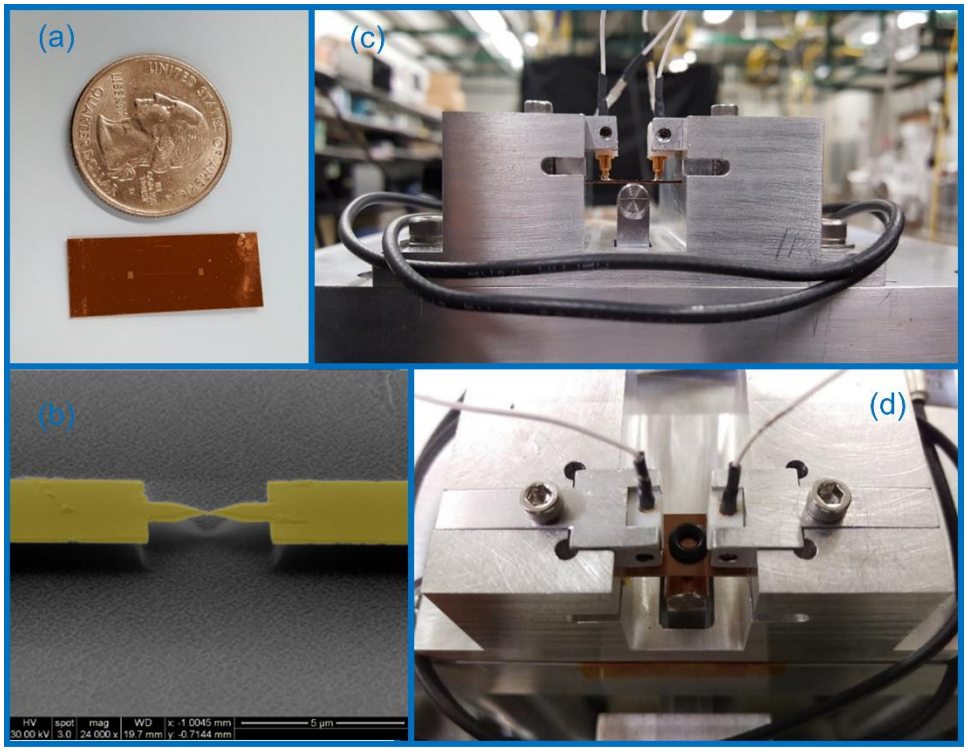

Figure S1. Images of MCBJ set-up. (a) Example of a lithographically defined MCBJ sample. (b) False color SEM image at 45° showing the suspended gold bridge in the center of a sample. (c) Side view of bending apparatus showing clamped-in sample with push rod underneath. (d) Top view of a clamped-in sample showing a Kalrez gasket placed around the center of the junction.

S.2 Inter-Electrode Distance Calibration

In an MCBJ set-up, the amount by which the two nano-electrodes pull apart (inter-electrode distance, Δx) for a given vertical movement of the push rod (piezo distance, Δz) is given by the "attenuation ratio"

($r = \Delta x/\Delta z$). While r can be calculated using a simple model of elastic bending, this result tends to be wrong by a factor of 2 to 4 due to the inhomogeneous elastic properties of real lithographically defined junctions.[1] It is therefore preferable to experimentally determine the attenuation ratio via one of several possible calibration methods.[2] Because the attenuation ratio depends on the exact length of the suspended gold bridge, which varies from sample to sample, we independently calibrated r for each sample considered in this work. To enable this calibration, each sample was run "bare" (no molecules deposited, only pure solvent which quickly evaporates) for a few thousand traces, and the r value calculated from these traces was then applied to all subsequent traces collected with that sample.

For the calibration itself, we employ the method of the tunneling slope. For small bias voltages, the tunneling current between two nano-electrodes as a function of their separation, x, is well-approximated by $I(x) = I_0 \exp(-Bx)$, where B is a constant depending on the effective work function of the electrodes.[3] A plot of $Log_{10}(G/G_0)$ vs. distance should therefore have a constant slope. By comparison to an STM-BJ set-up, Hong et al. found that this slope is 5.5 to 6 decades/nm for breaking traces collected in argon.[4] In an independent study, Grüter et al. found that the tunneling slope is ~1.7 times smaller for traces collected in air compared to in vacuum.[5] Based on high-quality data collected under vacuum,[2] this implies traces collected in air should have a tunneling slope of ~6 decades/nm, in agreement with the Hong et al. result, and thus we use this value for our calibration.

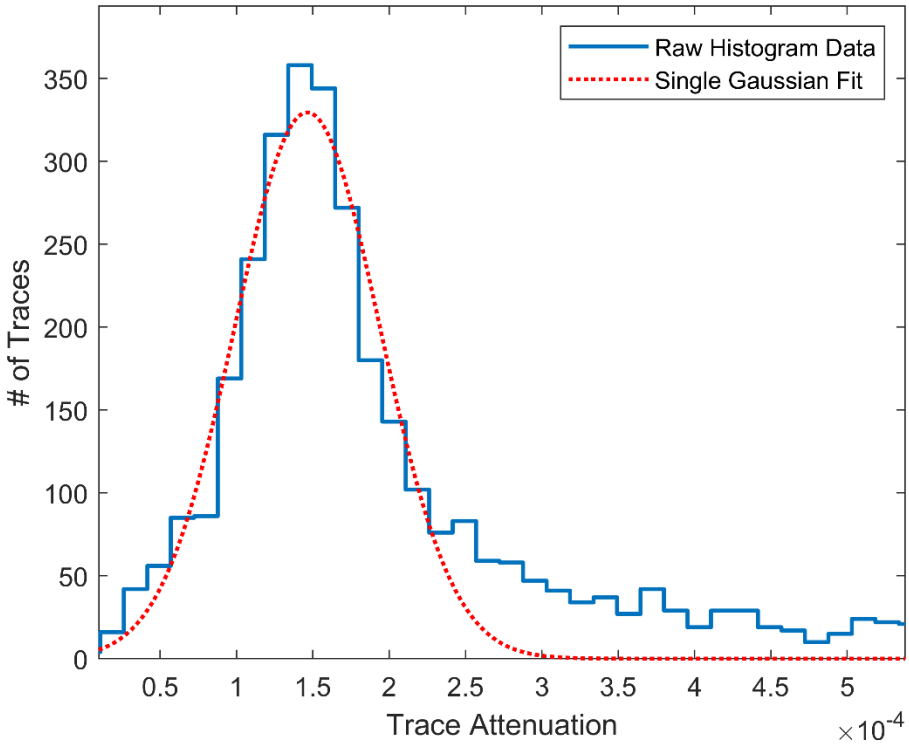

Figure S2. Histogram of 3481 individual trace attenuations (blue) and Gaussian fit (dotted red) used to determine the attenuation ratio for sample 113-2. Bin width was set to 1.54×10^{-5} using the Freedman–Diaconis rule.

To perform the calibration, we first linearly fit the portion of each breaking trace below 2×10^{-4} G_0, since this ensures that the tunneling slope is reliable,[2] and above 10^{-5} G_0, to be comfortably above the value of the noise floor of our amplifier.[6] We then calculated an attenuation ratio for each trace by assuming that the tunneling slope is 6 decades/nm. Next, a histogram of these attenuation ratios was constructed using the Freedman–Diaconis rule to determine the bin width, and finally we fit this histogram with a single unrestricted Gaussian (e.g. **Figure S2**). The peak of this Gaussian was taken as the attenuation ratio for all traces collected with the same sample. **Table S1** shows the number of tunneling traces used to calibrate each sample considered in this work and the resulting attenuation ratios.

Table S1. List of tunneling datasets considered in this work. One such tunneling dataset was collected for each MCBJ sample after depositing pure solvent (which quickly evaporates) but before depositing molecules. These tunneling datasets were used to determine an attenuation ratio for each sample, which was then applied to all subsequent datasets collected using that same sample.

Dataset ID#	Sample #	Molecule Later Deposited on Sample	# of Traces	Attenuation Ratio/10^{-4}	Solvent Used
135	108-4	C6-2SMe	6460	1.01	Hexanes
127	111-4	OPV2-2BT	1537	1.02	Dichloromethane
130	108-5	OPV2-2BT	3847	1.52	Dichloromethane
49	097-2	OPV3-BT-Br	3469	2.01	Dichloromethane
82	106-1	OPV3-BT-Br	5777	1.96	Dichloromethane
85	098-4	OPV3-BT-Cl	2148	1.18	Dichloromethane
102	102-5	OPV3-BT-Cl	3881	2.30	Dichloromethane
105	101-4	OPV3-BT-CN	7664	1.58	Dichloromethane
112	101-3	OPV3-BT-CN	2084	1.55	Dichloromethane
117	114-2	OPV3-BT-CN	2206	1.42	Dichloromethane
120	103-2	OPV3-BT-CN	8009	1.98	Dichloromethane
30	098-2	OPV3-BT-F	4113	1.36	Dichloromethane
33	099-5	OPV3-BT-F	4894	1.47	Dichloromethane
94	099-1	OPV3-BT-F	4580	2.03	Dichloromethane
97	098-3	OPV3-BT-F	3454	1.72	Dichloromethane
1	113-2	OPV3-BT-H	3481	1.47	Dichloromethane
11	104-5	OPV3-BT-H	7122	1.17	Dichloromethane
54	104-4	OPV3-BT-H	3567	1.84	Dichloromethane
23	097-1	OPV3-BT-MeO	2523	1.27	Dichloromethane
27	111-2	OPV3-BT-MeO	2480	0.96	Dichloromethane
18	113-3	OPV3-BT-NO2	6269	1.21	Dichloromethane
21	113-4	OPV3-BT-NO2	3497	1.59	Dichloromethane

S.3 Additional Design Criteria for Segment Clustering

S.3.1 Trace Starting and Ending Criteria. In the BUS segmentation process, the first segment of each trace is forced to start at the first data point. It is thus important to use consistent starting criteria for every trace in a dataset to avoid any influence from confounding variables. For this work, we begin each trace the last time it passes below 2.5 G_0, to avoid issues with limited discrimination and accuracy of our amplifier at higher conductance values.[6] Modifications to these starting criteria do not meaningfully affect our results (see section S.5.3).

The ending criteria for each trace are similarly important. We first exclude any conductances below the noise floor of our amplifier[6] (typically 10^{-6} G_0, but slightly higher in a few datasets; see section S.4 for details). Additionally, in cases where a trace drops below the noise floor but then later returns to a higher conductance, we end the trace the *first* time it drops below this level. This is necessary to avoid large empty gaps in traces, since BUS is not designed to work in such cases.

S.3.2 Parameter Standardization. Standardizing the distribution of a variable typically involves dividing by the sample standard deviation. However, because the standard deviation is sensitive to outliers, this can skew the standardization process. In contrast, the range of the middle 80% of values in a dataset is quite insensitive to outliers, so we use this measure of spread to standardize the first three segment parameters (X_0, Y_0, and Log(L)). Because the θ and R^2 parameters have limited possible ranges—(-90° to 90°) and (0 to 1), respectively—we standardize them by dividing by 80% of those full possible ranges. This has the advantage of making the standardization process less dependent on a particular dataset.

Finally, θ is only calculated *after* the inter-electrode distance and log(conductance) dimensions have already been standardized. This is necessary to ensure that θ is fully independent of the units on the x- and y-axes.

S.3.3 Assignment of SOPTICS Parameter Values. The c_L and c_P SOPTICS parameters control how many random projections are performed, with larger values leading to a more stable and accurate approximation of the original OPTICS algorithm. The creators of SOPTICS found that $c_L = c_P = 20$ produced excellent results;[7] out of an abundance of caution, we use a higher value of $c_L = c_P = 30$ (see section S.5.1 for details).

The *minSize* parameter helps control how the random projections are sampled to find nearby points.[7] Because SOPTICS is extremely insensitive to the value of *minSize* over a large range (see section S.5.2), we fix its value at 120.

S.3.4 Length-Weighting of Segments. Because OPTICS/SOPTICS is a density-based clustering algorithm, the density of segment parameters in 5-dimensional space ultimately determines how segments are clustered, with the densest regions forming the "cores" of key clusters. However, because segments are drawn from traces of roughly the same length, there will almost always be many more short segments than long ones. Perversely, this leads to a *lower* density of long segments, even though they represent many *more* data points from the original traces, making it difficult to form clusters of long segments. To remedy this issue, in the density calculations we apply a weighting factor to each segment that is proportional to its length. This ensures that the density of segments in parameter space corresponds to the density of raw data points rather than the number of segments themselves. In practice, this weighting is accomplished by duplicating each segment in proportion to its length before clustering. This step introduces another parameter, *len_per_dup*, that controls how many times a segment of a given length is duplicated. This parameter also serves as the minimum segment length, as we exclude segments that are not long enough for even a single duplicate. We set *len_per_dup* to 0.05 nm (*e.g.*, segments between 0.20 and 0.24 nm long will have 4 total copies) to ensure that all segments down to the length of a single bond will be included. We also note that the effects of changing *len_per_dup* are correlated with the effects of changing *minPts*, the parameter that defines how density is estimated (see section S.5.4). Therefore, because we use 12 different values of *minPts*, we are already capturing much of the possible variation from using different values of *len_per_dup*. Segment duplication is performed after the parameterization step.

S.4 Dataset Collection and Construction

Pausing a sample to re-deposit molecular solution often leads to a discrete, qualitative change in trace behavior—e.g., the fraction of traces displaying a molecular plateau (the "molecular yield") may significantly increase or decrease after re-deposition, or the gold electrodes may undergo rearrangement,

as evidenced by a significant change in where bridge rupture occurs on the absolute push rod movement scale. Such changes may also occur when depositing pure solvent on a junction already containing molecules, or when "starting a new trial" by fully relaxing the push rod and junction, followed by restarting trace-collection. We therefore treat the traces collected during each deposition/trial combination for a given MCBJ sample as a separate dataset. In the context of clustering, splitting each sample into multiple datasets in this way is the conservative approach; if instead we clustered the traces from each sample as one big dataset, we would be much more likely to find "consistent" features because the algorithm might only identify the regions where multiple disparate features all overlap. Splitting datasets at the natural points where qualitative changes tend to occur challenges Segment Clustering by providing the most opportunities for it to be confounded by changes in the "background".

For this work, we did not consider datasets from samples which showed strong signs of contamination in their initial "pure tunneling" sections. We also excluded molecular datasets in which no molecular feature was apparent, insufficient traces were collected (significantly less than 1000), or obvious noise features were present. For the OPV3-2BT-X family, this left us with 43 different molecular datasets, each corresponding to an entire deposition/trial block of traces (**Table S2**). We observed no apparent correlation between the number of depositions or trials and junction conductance. In nearly all of these datasets, the noise floor was set to 10^{-6} G_0, the nominal bottom end of the range for our amplifier.[6] However, due to differences in calibration, a few samples displayed higher noise levels, requiring us to manually set a higher noise floor to prevent physically meaningless data from affecting clustering results (see **Table S2**).

Table S2. List of all OPV3-2BT-X datasets considered in this work. Each dataset corresponds to a full deposition/trial block of traces. All molecular solutions were 1 μM. The top-to-bottom order of datasets in this table corresponds with the left-to-right order of points in Figure 7.

Dataset ID#	Sample #	Trial #	Deposition #	# of Traces	Molecule Name	Noise Floor (G_0)
2	113-2	1	1	5424	OPV3-BT-H	1.0E-06
3	113-2	1	2	9446	OPV3-BT-H	1.0E-06
12	104-5	1	1	3545	OPV3-BT-H	1.0E-06
13	104-5	1	2	4550	OPV3-BT-H	1.0E-06
14	104-5	2	2	2997	OPV3-BT-H	1.0E-06
15	104-5	3	2	6280	OPV3-BT-H	1.0E-06
16	104-5	4	2	5062	OPV3-BT-H	1.0E-06
58*	104-4	2	3	4113	OPV3-BT-H	1.0E-06
59	104-4	3	3	6294	OPV3-BT-H	1.0E-06
25	097-1	2	1	4065	OPV3-BT-MeO	1.0E-06
26	097-1	2	2	3137	OPV3-BT-MeO	1.0E-06
28	111-2	2	1	4051	OPV3-BT-MeO	1.0E-06
29	111-2	2	2	6214	OPV3-BT-MeO	1.0E-06
31	098-2	1	1	5182	OPV3-BT-F	1.0E-06
95	099-1	1	1	7695	OPV3-BT-F	1.0E-06
96	099-1	1	2	2147	OPV3-BT-F	1.0E-06
34	099-5	1	1	7922	OPV3-BT-F	1.0E-06
35	099-5	1	2	18568	OPV3-BT-F	1.0E-06
37	099-5	2	3	3941	OPV3-BT-F	1.0E-06
98	098-3	1	2	8661	OPV3-BT-F	1.0E-06
99	098-3	1	4	8753	OPV3-BT-F	1.0E-06
101	098-3	2	5	4120	OPV3-BT-F	1.0E-06
86	098-4	2	1	2940	OPV3-BT-Cl	1.0E-06
88	098-4	3	2	7670	OPV3-BT-Cl	1.0E-06
103	102-5	1	1	6394	OPV3-BT-Cl	1.0E-06
104	102-5	1	2	7841	OPV3-BT-Cl	1.0E-06
50	097-2	1	1	8603	OPV3-BT-Br	1.0E-06
51	097-2	1	2	10529	OPV3-BT-Br	1.0E-06
83	106-1	1	1	9572	OPV3-BT-Br	1.0E-06
84	106-1	1	2	15707	OPV3-BT-Br	1.0E-06
19	113-3	1	1	7310	OPV3-BT-NO2	1.0E-06
20	113-3	1	2	8083	OPV3-BT-NO2	1.0E-06
22	113-4	1	1	7799	OPV3-BT-NO2	1.0E-06
107	101-4	2	2	6679	OPV3-BT-CN	5.5E-06
108	101-4	2	3	7449	OPV3-BT-CN	5.5E-06
109	101-4	2	4	2309	OPV3-BT-CN	5.5E-06
114*	101-3	1	2	2772	OPV3-BT-CN	1.0E-05
116	101-3	2	3	5477	OPV3-BT-CN	1.0E-05
118	114-2	1	3	4280	OPV3-BT-CN	3.0E-06
121	103-2	1	1	10259	OPV3-BT-CN	1.0E-06
123	103-2	2	2	3175	OPV3-BT-CN	1.0E-06
125	103-2	3	3	2783	OPV3-BT-CN	1.0E-06
126**	103-2	3	3	6548	OPV3-BT-CN	1.0E-06

*Dataset not included in analysis; see section S.7 for details.

**Pure dichloromethane was deposited between datasets #125 and #126; hence they are treated as distinct datasets even though they have the same trial number and number of molecular depositions.

For this work, we also considered five OPV2-2BT datasets and two C6-2SMe datasets (**Table S3**). In two of these cases, the dataset consisted of a subset of consecutive traces from a deposition/trial block in order to exclude clear noise features (see **Table S3**). We then constructed eight different 1:1 synthetic mixtures of these OPV2-2BT and C6-2SMe datasets by combining different sets of traces from different datasets. Because the OPV2-2BT datasets contained more traces, for each mixture we used *all* of the traces from one of the C6-2SMe datasets and then added an equivalent number of consecutive traces from a subset of one of the OPV2-2BT datasets (see **Table S4** for details).

Table S3. List of all OPV2-2BT and C6-2SMe datasets considered in this work. "Subset" refers to datasets corresponding to a consecutive subset of traces from an entire deposition/trial block, taken to exclude clear noise features. All noise floors are 10^{-6} G_0.

Dataset ID#	Sample #	Trial #	Deposition #	Subset	# of Traces	Molecule Name	Solution Concentration (μM)	Solvent Used
128	111-4	1	1	No	3234	OPV2-2BT	1	Dichloromethane
129	111-4	1	2	No	2680	OPV2-2BT	1	Dichloromethane
132	108-5	1	2	Yes	2400	OPV2-2BT	1	Dichloromethane
133	108-5	3	5	No	6562	OPV2-2BT	1	Dichloromethane
134	108-5	4	1*	No	5807	OPV2-2BT	10	Dichloromethane
136	108-4	1	2	Yes	1315	C6-2SMe	10	Hexanes
137	108-4	2	2	No	1065	C6-2SMe	10	Hexanes

*1st deposition of a 10 μM solution, but 6th depositon overall (first 5 depositions were each with a 1 μM solution).

Since each sample has a slightly different attenuation ratio, the density of data points on the inter-electrode distance scale is also different for each sample. This is an issue for constructing synthetic mixture datasets because it would cause the denser dataset to have extra weight in what is supposed to be a 1:1 mixture. We therefore used linear interpolation to resample all OPV2-2BT and C6-2SMe traces at a rate of one data point per 4×10^{-4} nm of inter-electrode distance. This resampling was performed before clustering the pure datasets and before the construction and clustering of the synthetic mixture datasets.

Table S4. List of the eight different 1:1 OPV2-2BT:C6-2SMe synthetic mixture datasets created for this work, along with details of their construction. Dataset ID #s refer to **Table S3**. Mixture #1 is the dataset used for Figure 8g-i.

Mixture #	Total # of Traces	Dataset ID for OPV2-2BT Traces	Dataset ID for 2,9-dithiadecane	Traces Used from OPV2-2BT Dataset
1	2630	134	136	1-1315
2	2130	134	137	1-1065
3	2630	133	136	1-1315
4	2630	128	136	1-1315
5	2630	132	136	1-1315
6	2130	132	137	1-1065
7	2630	129	136	1-1315
8	2630	134	136	1500-2814

S.5 Robustness of OPV3-2BT-X Results to Clustering Parameters

S.5.1 Robustness to Random Seed. The SOPTICS algorithm employs random projections in order to achieve its improved clustering times, and even regular OPTICS, when properly implemented, uses a random choice for the first point in the cluster order. If the clustering structure extracted by these algorithms is truly inherent to the dataset, then the clustering results should not be meaningfully affected by using a different set of random numbers. To confirm that this is the case for our OPV3-2BT-X results, we re-clustered one of our datasets (ID# 3 in **Table S2**) using ten different random seeds for MatLab's pseudo-random number generator. This is also a good way to evaluate our choice for the parameters c_L and c_P; because these parameters control how many different random projections are used by SOPTICS, we know that their values are suitably large when the clustering outputs for different random seeds all converge to give the same results. We therefore repeated this random seed testing for three different values of $c_L = c_P$. For this testing we fixed the value of *minPts* at 85.

We used two different methods to evaluate the similarity of these different clustering results. First, we simply compared the peak conductance value for the main plateau cluster in this dataset, as this peak conductance is what we are ultimately interested in for our analysis of the OPV3-2BT-X family. Second,

we used a similarity index developed by Rand to compare the entire clustering *solutions* that each main plateau cluster belongs to. The Rand similarity index is a pairwise comparison that ranges from 0 to 1, with 1 meaning that every data point was assigned to the same cluster in both clustering solutions and 0 meaning that every data point was assigned to a different cluster in one solution vs. the other.[8] Because this method compares the overall clustering structure instead of just the peak value of a single cluster, it provides a more stringent test of the similarity of different clustering results.

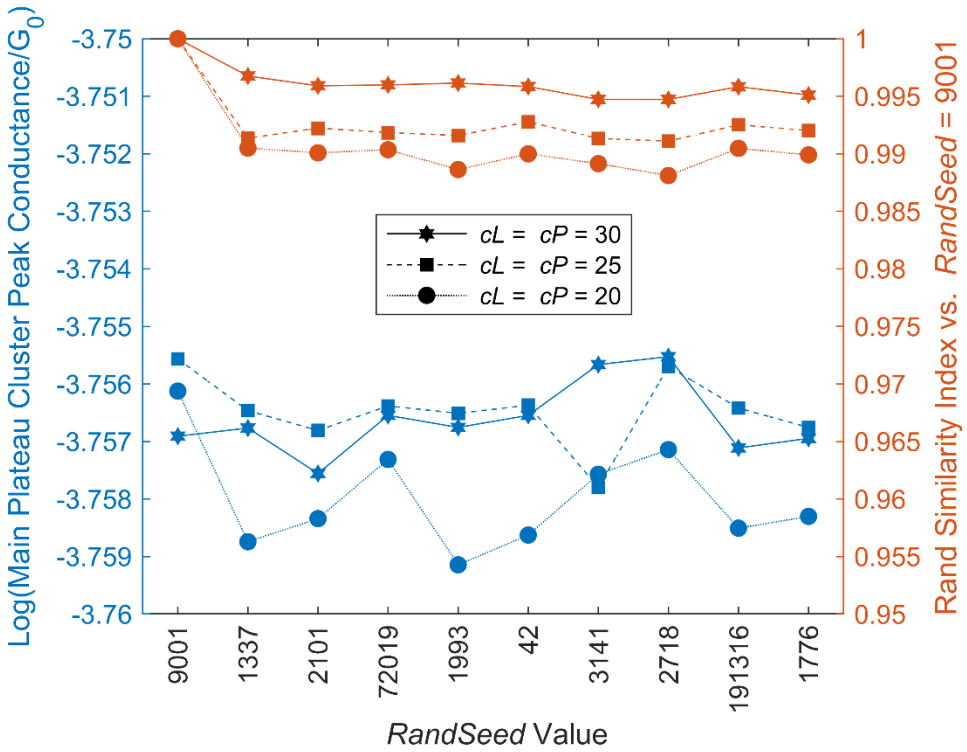

Figure S3. Comparison of fitted peak conductance values for the main plateau cluster for a single OPV3-2BT-H dataset clustered using 10 different random seeds and three different values for the parameters *cL* = *cP*, with the *minPts* parameter fixed at 85 (left axis). For the right axis, the clustering *solution* which contained the main plateau cluster for each of the 30 clustering outputs was first identified. Each of these solutions was then compared to the solution for a random seed of 9001 using the Rand similarity index. These results demonstrate both that SOPTICS is not affected meaningfully by random seed choice and that *cL* = *cP* is set to a sufficiently large value.

The results of these evaluation methods for our random seed testing are summarized in **Figure S3**. Even for $c_L = c_P = 20$, changing the random seed has essentially no effect, with the conductance peak varying

by less than 0.003 decades and the Rand Similarity Index always greater than 0.985. Our decision to use $c_L = c_P = 30$, where the convergence is even tighter, is thus clearly a very safe choice.

In addition, these results demonstrate that in our implementation SOPTICS is essentially unaffected by the set of random numbers used, and is thus behaving properly. For the clustering results discussed in the main body of the paper and for all subsequent testing, we therefore used the last digits of the system time to generate a different random seed for each clustering run.

S.5.2 Robustness to *minSize*. To ensure that our OPV3-2BT-X results are not dependent upon our choice for the *minSize* parameter, we re-clustered another dataset (ID# 25 in **Table S2**) using 17 different values of *minSize*. We again fixed the value of *minPts* at 85 for this testing.

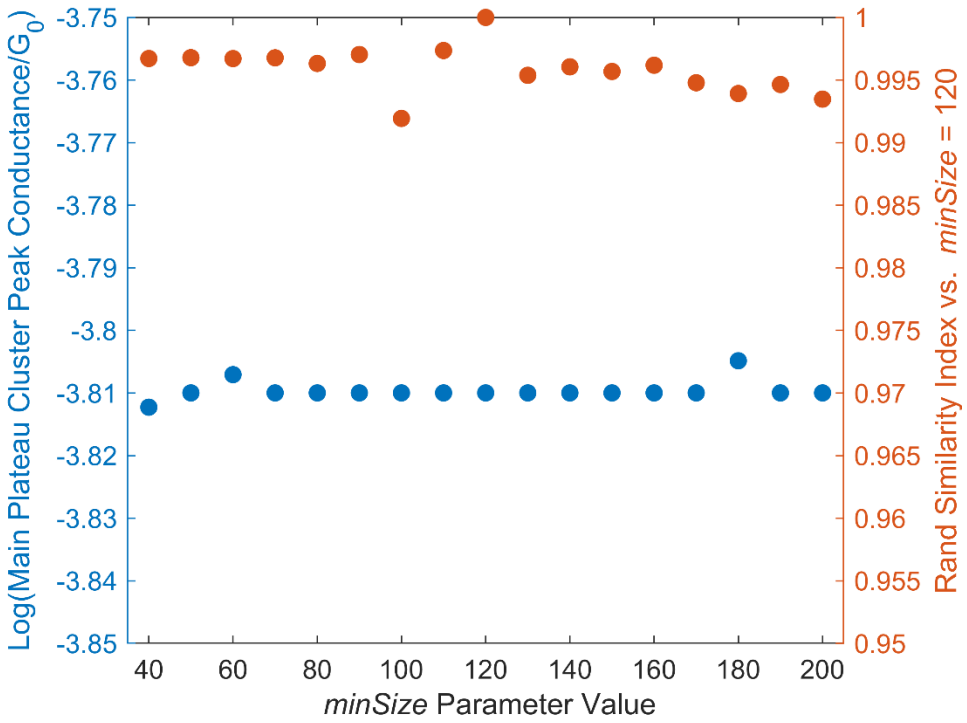

Figure S4. Comparison of fitted peak conductance values for the main plateau cluster for a single OPV3-2BT-MeO dataset clustered using 17 different values for the parameter *minSize*, with the *minPts* parameter fixed at 85 (left axis). For the right axis, the clustering *solution* which contained the main plateau cluster for each of the 17 clustering outputs was first identified. Each of these solutions was then compared to the solution for *minSize* = 120 using the Rand similarity index. These results demonstrate that the exact value of *minSize* is not very important for the behavior of SOPTICS, and so it is safe to use a single fixed value for this parameter.

We used the same two evaluation methods (main plateau cluster peak conductance and Rand similarity index) described in section S.5.1 to compare these different clustering results. As shown in **Figure S4**, the clustering output is extremely insensitive to the choice of *minSize* over quite a large range. This justifies our choice to fix the value of *minSize* at 120.

S.5.3 Robustness to Trace Starting Criteria. As described in section S.3.1, to ensure consistent starting criteria before the segmentation step, we begin each trace the last time it passed below a conductance of 2.5 G_0. To check that our OPV3-2BT-X results do not depend on this choice, we re-clustered another of our datasets (ID# 19 in **Table S2**) using six different values for this "*TopChop*" conductance value.

Because changing the *TopChop* affects the segmentation step, these different clustering outputs do not contain the exact same objects for clustering, and so cannot be compared using the Rand similarity index. However, comparing the peak conductance of the main plateau cluster for each of these results (**Figure S5a**) shows that the choice of the *TopChop* value does not meaningfully impact our results.

As an additional test, we also considered a different type of starting criteria: instead of a "*TopChop*", a "*LeftChop*" in which we begin each trace at zero inter-electrode distance. Comparing the results for six of our datasets for these two different chop methods (**Figure S5b**) again confirms that our OPV3-2BT-X conclusions are not dependent upon our choice of starting criteria. We note that this left chop at zero significantly improves clustering time by reducing the number of data points, and so may be preferred in some situations.

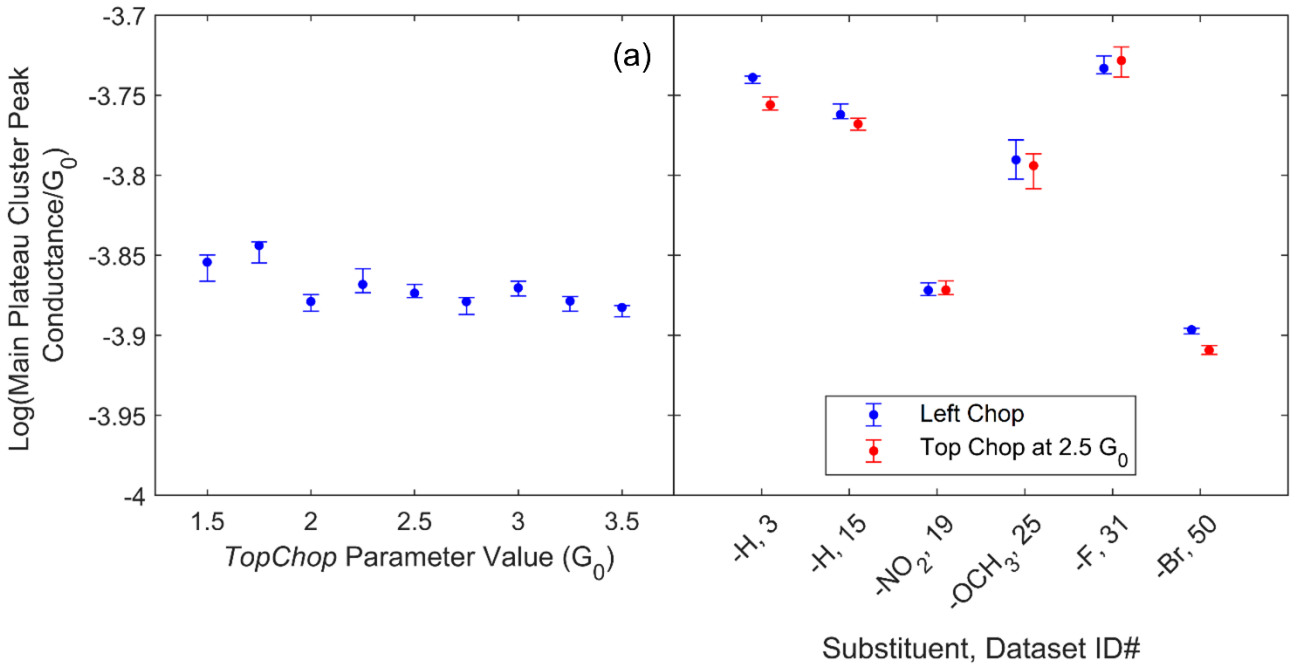

Figure S5. Demonstration of the insensitivity of OPV3-2BT-X clustering results to trace starting criteria. (a) Peak conductance values for the main plateau cluster of the same OPV3-2BT-NO_2 dataset clustered using 9 different "*TopChop*" values (only the portion of each trace after the last time its conductance passes below *TopChop* is included for clustering). (b) Comparison of the peak conductance values for the main plateau clusters for six different OPV3-2BT-X datasets (dataset ID#s refer to **Table S2**) clustered using a *TopChop* of 2.5 G_0 (red) or a "*LeftChop*" (blue), in which only the portion of each trace after zero inter-electrode distance is included for clustering.

S.5.4 Robustness to *len_per_dup* and Correlation with *minPts*. As described in section S.3.4, the parameter *len_per_dup* controls how often each segment is duplicated in proportion to its length (and also sets the minimum segment length). Decreasing *len_per_dup* increases the density of data points in all regions, and is thus expected to have a similar effect to decreasing the value of *minPts*. To confirm this, we re-clustered one of our datasets (ID# 2 in **Table S2**) at a variety of combinations of *minPts* and *len_per_dup* parameter values. The clustering solutions containing the main plateau cluster were then compared using the Rand similarity index as well as the peak plateau conductance (**Figure S6a,b**), as described in section S.5.1. Because *len_per_dup* also controls the minimum segment length, clustering runs with larger *len_per_dup* values used slightly fewer segments for clustering. Therefore, for each

pairwise comparison only those segments present in *both* clustering results were considered when computing the Rand similarity index.

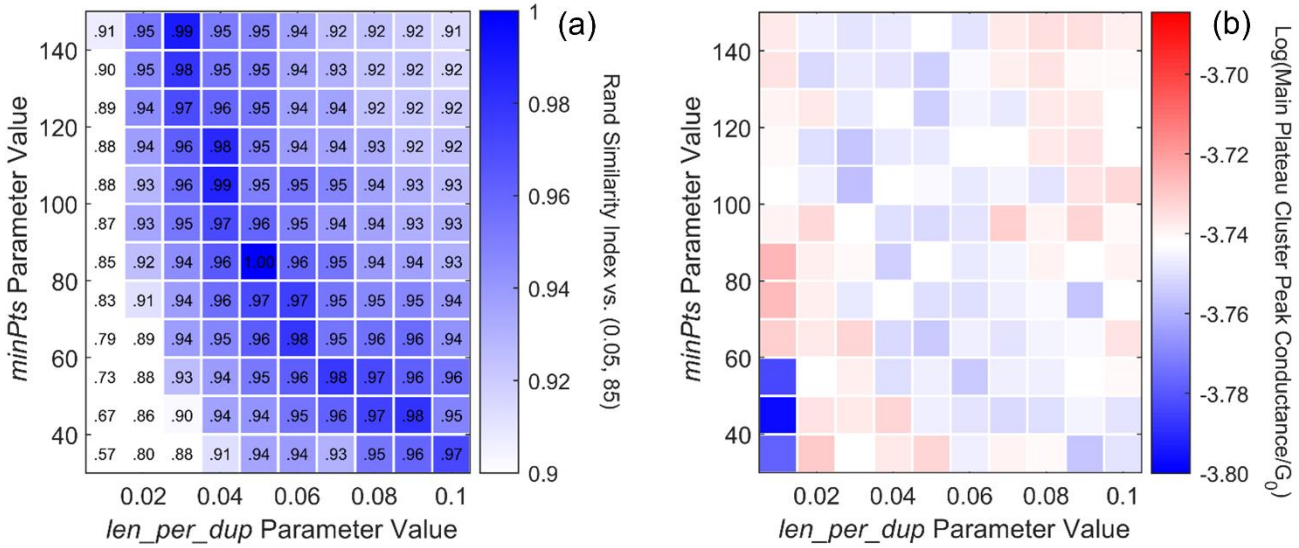

Figure S6. Comparison of outputs for a single OPV3-2BT-H dataset clustered using 120 different combinations of the *minPts* and *len_per_dup* parameters. (a) Rand similarity index for the clustering solution from each output which contained the main plateau cluster, compared to the chosen solution for the *minPts* = 85 and *len_per_dup* = 0.05 nm output. The fact that most of the index values are close to one shows that the clustering is relatively insensitive to these two parameters, and the northwest-to-southeast "ridge" demonstrates that they are positively correlated with each other. (b) Fitted peak conductance values for the main plateau cluster for each output, demonstrating that this measurement is quite insensitive to both parameters.

The high Rand similarity indices (**Figure S6a**) and similar peak conductance values (**Figure S6b**) that are found across a wide range of *len_per_dup* values indicate that clustering results are quite robust to changes in this parameter. More importantly, however, **Figure S6a** demonstrates that there is indeed a strong correlation between the effects of changing the *len_per_dup* and *minPts* parameters, as expected. This helps justify our decision to fix the value of *len_per_dup*, because it means that by using multiple values of *minPts* we are already capturing much of the variation that would be caused by changes to *len_per_dup*.

S.5.5 Robustness to Settings of Iterative L-Method. One of the advantages to using the Iterative L-Method as a stopping criterion for Bottom-Up Segmentation is that it is described as being parameter-free.

However, the algorithm does rely on a value, *minimum_cutoff_size*, which the authors argue can be considered a constant instead of a parameter because a value of 20 yields good results in a wide variety of contexts.[9] Out of an abundance of caution, we also tried re-clustering a handful of our datasets using a smaller (16) or larger (24) value of *minimum_cutoff_size*.

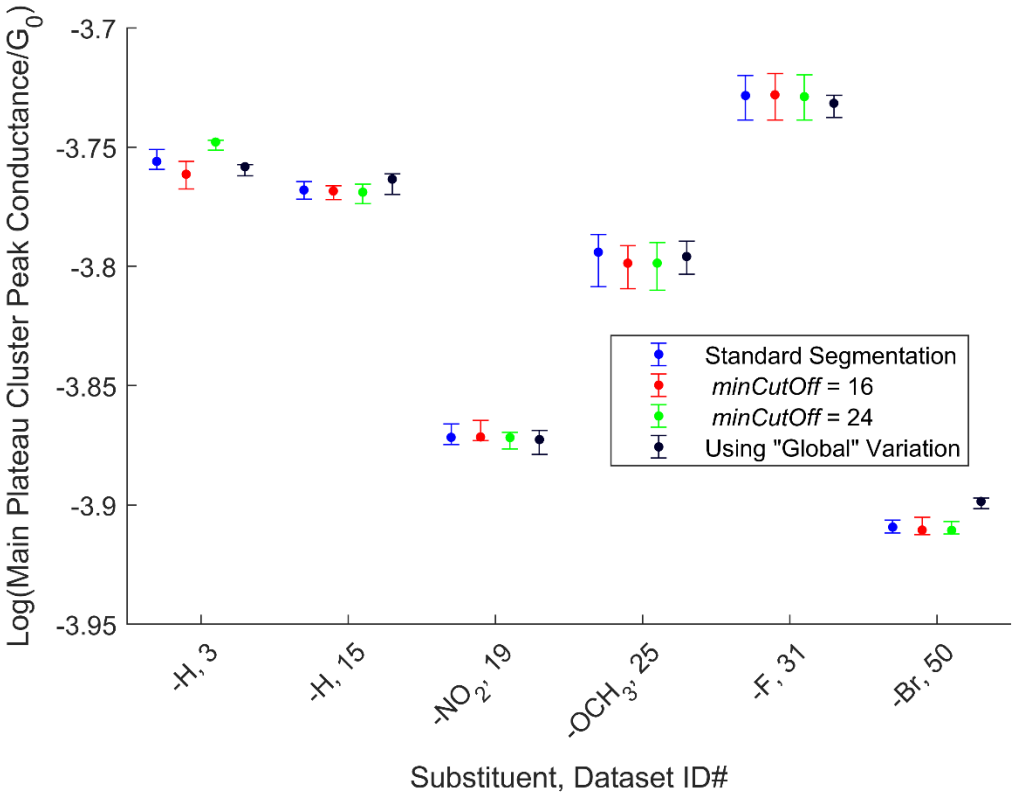

Figure S7. Comparison of the peak conductance values for the main plateau clusters for six different OPV3-2BT-X datasets (dataset ID#s refer to **Table S2**) clustered after using the "standard" segmentation procedure (blue); after segmentation with the *minimum_cutoff_size* value set to 16 (red) or 24 (green) instead of its standard value of 20; and after using the "Global" instead of the "Greedy" Iterative L-Method as stopping criteria for segmentation (black). These results demonstrate that slight variations in how the segmentation algorithm is implemented do not meaningfully affect our OPV3-2BT-X results.

Additionally, the authors actually present two slight variations of the Iterative L-Method: "Global" and "Greedy". As mentioned above, we use the "Greedy" Iterative L-Method because it was generally found to produce superior results.[9] However, again out of an abundance of caution, we also tried re-clustering these same datasets using the "global" Iterative L-Method instead. As shown in **Figure S7**, neither the

changes to *minimum_cutoff_size* nor the switch from "Greedy" to "Global" meaningfully affect our results for the OPV3-2BT-X molecules.

S.6 Selecting Clusters from Multiple Cluster Outputs for the Same Dataset

As discussed in the main text, each dataset in this work was re-clustered twelve times using different values of the parameter *minPts* in order to account for uncertainty in the "optimal" setting for this parameter. For the figures in this work, we calculated and show each clustering output for *minPts* = 85 (roughly in the center of the 12 different *minPts* values).

After selecting a particular full-valley cluster of interest in the *minPts* = 85 output of a given dataset (e.g. the main plateau cluster for each OPV3-2BT-X dataset), we employed an automated algorithm to identify the analogous full-valley cluster in each of the other eleven clustering outputs for that same dataset. This algorithm first calculates the median value of each normalized segment parameter for the manually chosen cluster as well as for every full-valley cluster in the other eleven outputs. It then selects the single full-valley cluster from each of those outputs with the smallest Euclidean distance between its "median centroid" and that of the manually chosen cluster. The clusters identified with this automated algorithm matched the unambiguous assignments that would have been made by eye.

When the distributions for chosen clusters were fit to determine peak conductance values, the clusters from the twelve different outputs for each dataset were fit independently to obtain twelve different peak values. To represent the peak conductance of a single dataset (specifically, in Figure 7, **Figure S5**, **Figure S7**, and **Figure S17**), we use the median from among these twelve peak values, along with error bars representing the range of the middle eight of the twelve values (i.e. the middle 66.7%).

S.7 Selection of Main Plateau Clusters for OPV3-2BT-X Datasets

Of the 43 OPV3-2BT-X datasets listed in **Table S2**, one dataset (ID# 114) did not produce any full-valley clusters that came close to corresponding to the molecular feature in the 2D histogram (possibly

because the percentage of junctions containing a molecule was too low), and so was excluded from subsequent analysis. In 31 cases, only a single full-valley cluster had any similarity to the molecular feature, and each of these clusters was quite similar to the main plateau cluster shown in Figure 4h. We therefore unambiguously assigned each of these clusters as the analogous "main plateau cluster" for their respective datasets.

In 10 of the OPV3-2BT-X datasets, two full-valley clusters were found which might correspond well to the molecular feature region in the 2D histogram. However, in each of these cases, one of the clusters consisted of mostly flat segments like the main plateau clusters in the 31 datasets mentioned above (*e.g.* **Figure S8a,d,g,j**), whereas the second cluster consisted of more angled segments at slightly higher conductance (*e.g.* **Figure S8b,e,h,k**). Moreover, the valley corresponding to each flatter cluster always showed up in a similar location in its reachability plot as the other identified main plateau clusters (*e.g.* **Figure S8c,f,i,l**), suggesting that it represents an analogous component of the dataset's hierarchical structure. Therefore, in these 10 datasets there was still a single unambiguous choice for which full-valley cluster was the analogous feature to the cluster in Figure 4h and should thus be assigned as the main plateau cluster. **Figure S8** compares the chosen main plateau clusters with the angled clusters for four examples from these 10 datasets to demonstrate how clear these choices were.

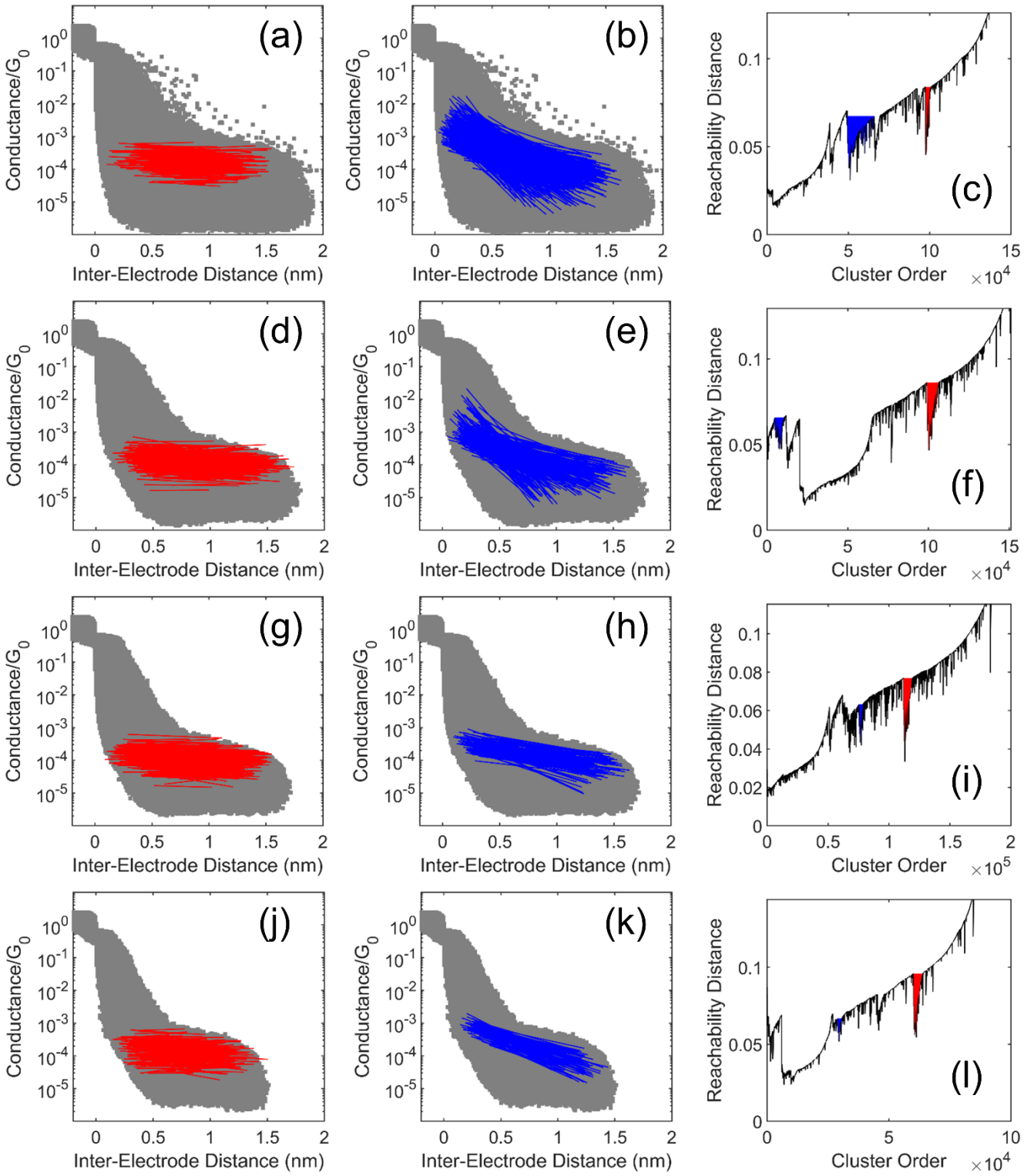

Figure S8. (a) Main plateau cluster chosen for dataset #59 (see **Table S2**). (b) Second full-valley cluster discovered in dataset #59 which corresponds well with the molecular feature from the 2D histogram, but is qualitatively distinct from the other identified main plateau clusters due to its higher conductance and more-angled segments. (c) Reachability plot for dataset #59 with the valleys corresponding to the clusters

in (a) and (b) highlighted, showing how they fit into the hierarchical clustering structure. (d-f) Analogous plots for dataset #103. (g-i) Analogous plots for the dataset #104. (j-l) Analogous plots for dataset #37. Together, these four examples demonstrate that even in the datasets containing multiple molecule-like full-valley clusters, there was consistently an unambiguous choice for which cluster was structurally most analogous to the cluster in Figure 4h and should thus be assigned as the main plateau cluster (*i.e.*, the flatter clusters in the first column).

Finally, in one OPV3-2BT-H dataset (ID# 58), only a single full-valley cluster corresponding to the molecular feature was found (**Figure S9**), but this cluster resembled the angled clusters discussed above much more than the main plateau clusters identified in the other 41 datasets. This is therefore the second OPV3-2BT-X dataset that we excluded from subsequent analysis because this cluster does not appear to belong in the same category as the other 41. No qualitative change to our conclusions would have resulted from inclusion of this dataset.

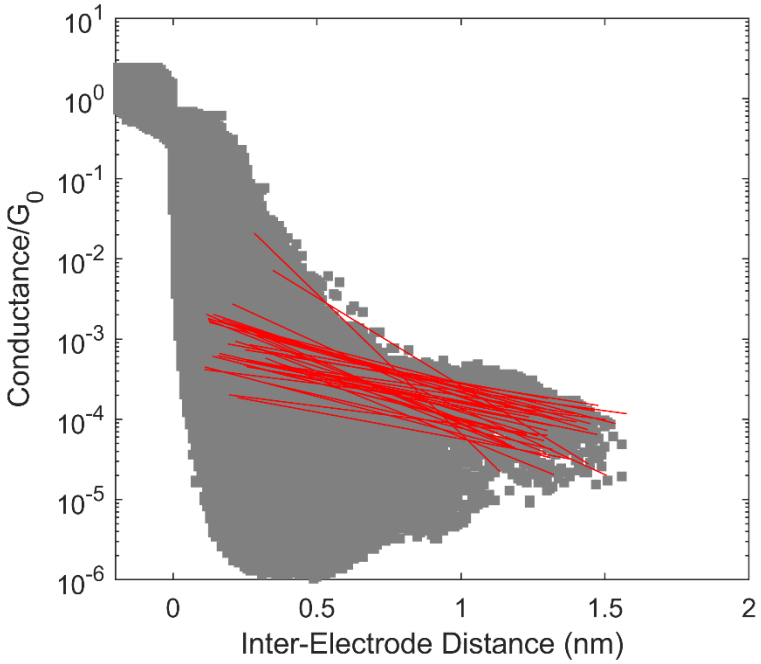

Figure S9. The only full-valley cluster from dataset #58 (see **Table S2**) which corresponds to the molecular feature in the 2D histogram. Because this feature seems to match the "secondary", angled clusters in **Figure S8** more than all other chosen main plateau clusters, it was excluded from subsequent analysis.

It is intriguing to note that the higher-conductance, more-angled clusters discovered in the 11 datasets discussed above appear qualitatively similar to the "class 2" traces identified by Cabosart et al. for a

structurally similar molecule using a completely different clustering approach. In an additional similarity, Cabosart et al. also found a lower-conductance, flatter cluster ("class 3" traces) which they assign to the "standard" binding configuration and find to be a consistent representation of the molecular conductance.[10] This perhaps suggests that these two features might be a conserved motif of rod-like conjugated molecules, and full atomistic calculations are needed to investigate this question in more detail. On a more general level, the fact that significantly different clustering methods identify similar molecular features supports the view that clustering analysis is an appropriate means of revealing intrinsic data structure.

S.8 Peak Fitting

In order to have a point of comparison to our main plateau cluster peak fits, we pursued the standard approach of fitting the molecular peak in each raw 1D histogram with a single Gaussian. However, due to the complex and asymmetric peak shape, fitting within the conductance range surrounding the molecular peak typically leads to unreasonable results (*e.g.* dotted green line in **Figure S10**), and moreover can strongly depend on exactly how this conductance range is defined. Therefore, to fit the raw 1D histogram molecular peaks for our OPV3-2BT-H datasets, we used an iterative approach to set the conductance bounds for fitting. Each histogram is first fit with a single Gaussian peak while only considering the conductance range -5.5 G_0 to -2.5 G_0 (*e.g.* the dotted green line in **Figure S10**). Ten more restricted fits are then performed, with the conductance bounds modified each time based on the results of the previous fit. At each iteration, the conductance bounds are centered around the peak value from the previous fit, and the width of this fitting region is 2 decades for the first two iterations, 1.5 decades for the next four, and 1 decade for the last four. This process was empirically found to produce reasonable fits for the eight OPV3-2BT-H datasets we applied it to (*e.g.* dashed red line in **Figure S10**), and the peak value always fully converged by the tenth iteration.

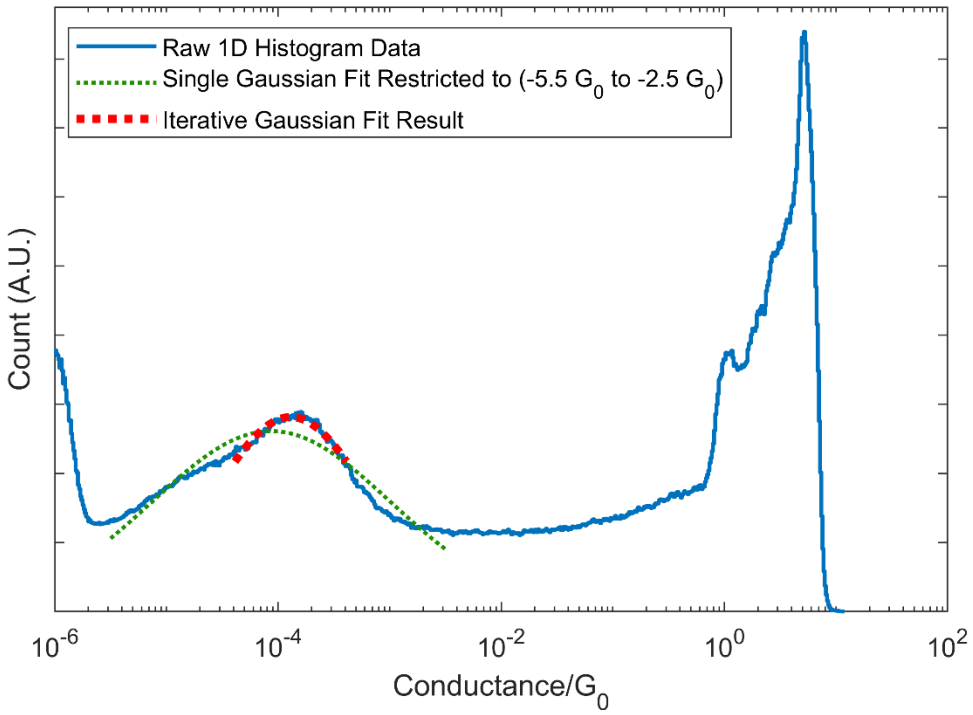

Figure S10. Raw 1D histogram for the OPV3-2BT-H dataset from Figure 1 (blue), along with a single Gaussian fit to only the range -5.5 G_0 to -2.5 G_0 (dotted green), and the result of an iterative process described in the text for determining the fitting range (dashed red).

For fitting the distributions of conductance values from specific clusters, in every case we used a single, unrestricted Gaussian fit. In the majority of cases, these distributions matched a Gaussian peak shape extremely well (e.g. Fig. 5). Some of the distributions displayed minor asymmetry or increased kurtosis, and thus fit a Gaussian peak shape less well; **Figure S11** shows the worst examples from the OPV3-2BT-X datasets. However, even in these cases, the single unrestricted Gaussian fit provided very reasonable approximations to the peaks and peak centers. A more complex fitting function would likely tighten the distributions of peak values in Figure 7; for example, adding a second fitting peak for the OPV3-2BT-Br and OPV3-2BT-Cl main plateau cluster distributions shown in **Figure S11c** and **Figure S11e**, respectively, would increase the conductance of the "main" peak, and these two datasets are both mild outliers on the low side in Figure 7.

For all histogram fitting in this work, the histogram bin width was determined based on the Freedman–Diaconis rule.

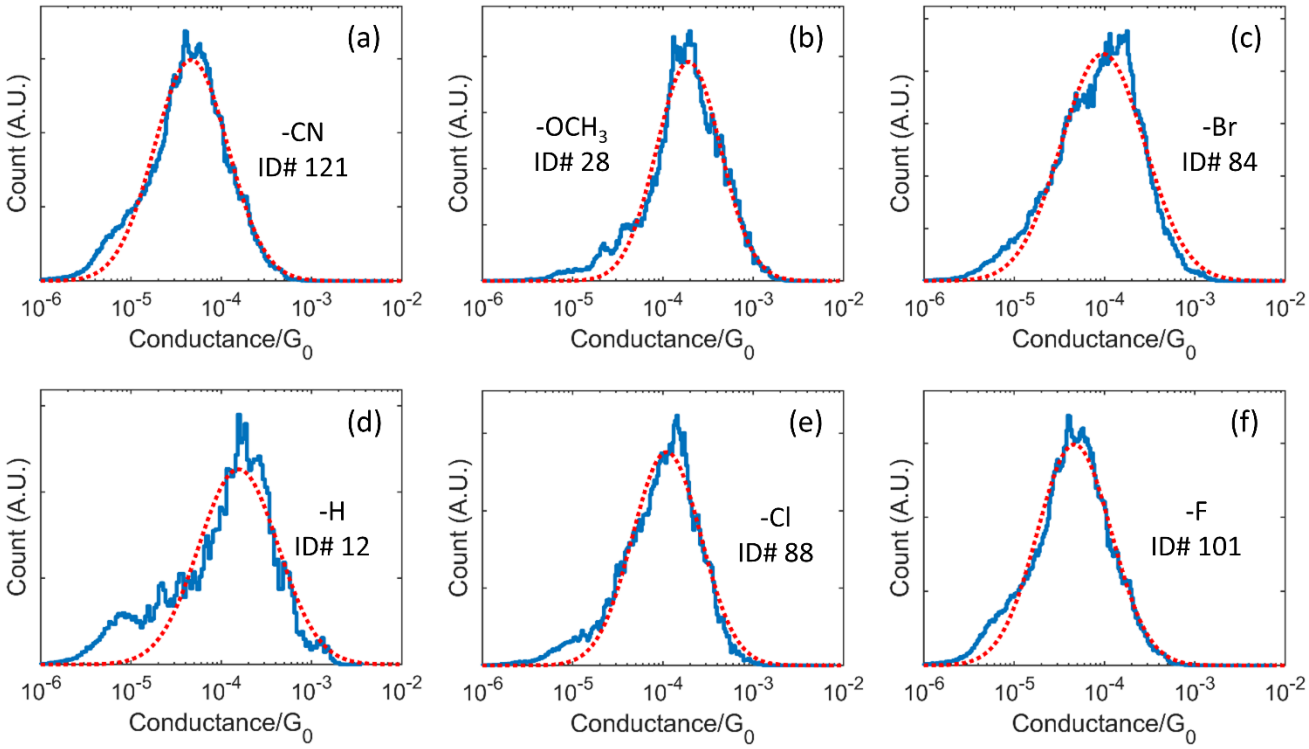

Figure S11. Main plateau cluster distributions (blue) and their respective unrestricted Gaussian fits (dotted red) for the six OPV3-2BT-X datasets in which these distributions were least Gaussian-shaped. The substituent, -X, and the ID# (from **Table S2**) for each dataset are inset for each plot.

S.9 Investigating OPV3-2BT-X Main Plateau Cluster Lengths

To help support our hypothesis that the main plateau cluster for each OPV3-2BT-X dataset represents the primary molecular feature, we investigated the maximum junction gap sizes implied by these clusters with two similar approaches. In the first method, we focus only on the actual trace pieces represented by the segments in the main plateau cluster. The end points of these trace pieces represent the maximum extent of each identified molecular plateau. However, it is possible that the linear features identified by Segment Clustering do not represent the entire time the molecule spent in the junction (e.g. the conductance may vary significantly during the detachment process). Therefore, in the second method we consider each entire trace containing a segment assigned to the main plateau cluster. The last time each

trace drops to a low value well below the conductance of the molecule (here the value of $5 \cdot 10^{-6}$ G_0 is used) is an alternative way to represent the distance at which the molecule fully breaks off. Both methods are demonstrated for an example OPV3-2BT-H dataset in **Figure S12a-d**.

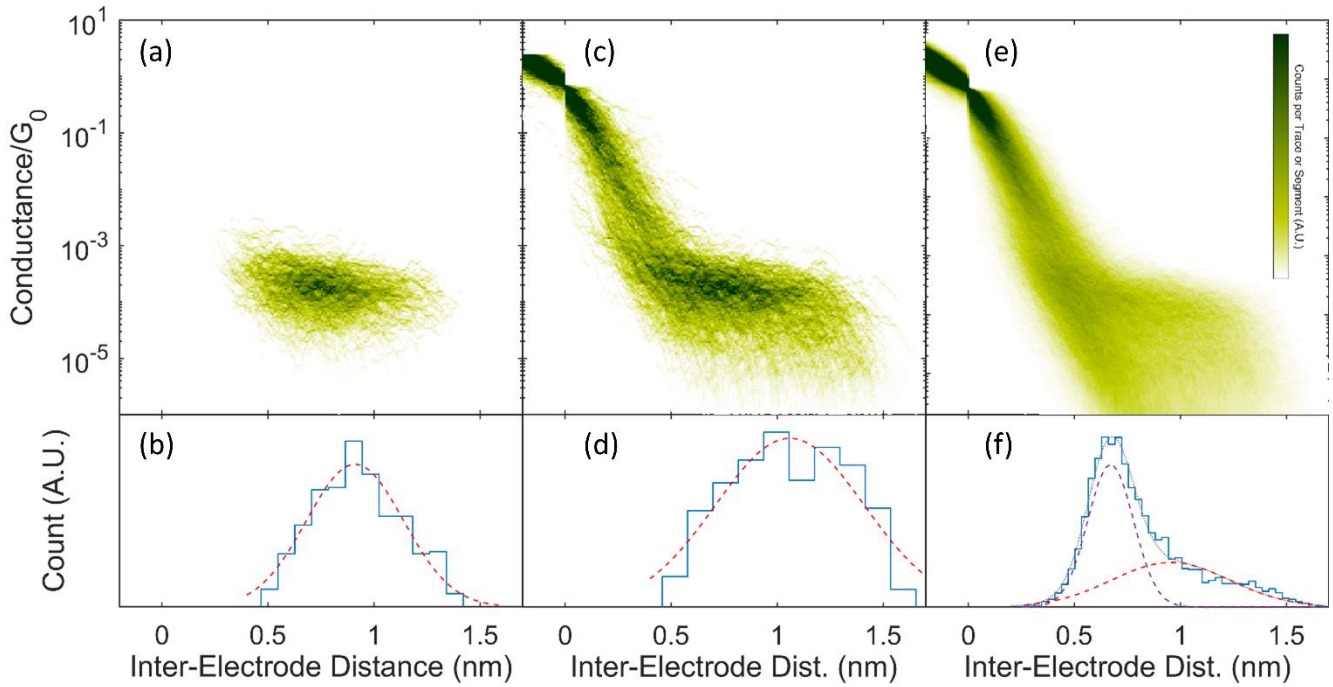

Figure S12. Examples of distance investigation methods using the OPV3-2BT-H dataset from Figure 1c. (a) 2D histogram of just the trace pieces whose linear segments were assigned to the main plateau cluster. (b) 1D histogram of the endpoints of the trace pieces in (a), fit with a single Gaussian peak (red). (c) 2D histogram of all traces containing segments which were assigned to the main plateau cluster. (d) 1D histogram of the distances at which each trace in (c) last crossed below the conductance value $5 \cdot 10^{-6}$ G_0, fit with a single Gaussian peak (red). (e) 2D histogram of *all* traces in the dataset. (f) Analogous to (d), but for the traces shown in (e); fit with two Gaussians (purple and red, total fit in gray).

For comparison, we also show the results of applying the "trace-cross" method to *all* traces in the dataset (**Figure S12e,f**). This entire-dataset distance distribution exhibits two peaks, typically attributed to the break-off of tunneling traces and to molecular traces respectively.[11–14] As shown in **Table S5**, both distance distributions for the main plateau cluster are quite similar to the second peak in the entire-dataset distribution, providing clear evidence that what we label the "main plateau cluster" corresponds to what is generally considered to be the "primary" molecular feature. Similar results were obtained for the other OPV3-2BT-X datasets considered in this work. The moderate variation that was observed between

datasets is likely due in large part to small systematic errors in attenuation ratios, and the overall pattern did not suggest any systematic differences in length between different substituents.

The fairly broad distributions seen in **Figure S12b,d** indicate that not all junctions reach the same degree of elongation before breaking off. The distribution peaks are somewhat shorter than what would be expected for fully-elongated molecular junctions, which is consistent with previous results for molecules with –BT linker groups.[15] This suggests that molecules with this linker group may in general not reach full extension.

Table S5. The peak and half-width at half-maximum (HWHM) values for the red Gaussian fits shown in **Figure S12** panels b, d, and f, respectively.

	Peak (nm)	HWHM (nm)
Segment End Points	0.91	0.27
Segment-Containing Trace Crosses	1.06	0.41
All Trace Crosses	0.95*	0.33*

*For the higher-distance of the two Gaussian fits (red in **Figure S12f**).

S.10 Selection of Main Plateau Clusters for OPV2-2BT and C6-2SMe

Figure S13 shows all of the full-valley clusters discovered in the OPV2-2BT dataset from Figure 8d. The cluster in **Figure S13i** can be unambiguously chosen as the main plateau cluster for the high-conductance feature. None of the full-valley clusters corresponds well to the low-conductance feature in this dataset (the cluster in **Figure S13f** is the closest, but does not align well with the low-conductance feature on either axis in the 2D histogram). Similar main plateau clusters were identified in the other four OPV2-2BT datasets considered in this work (**Table S3**).

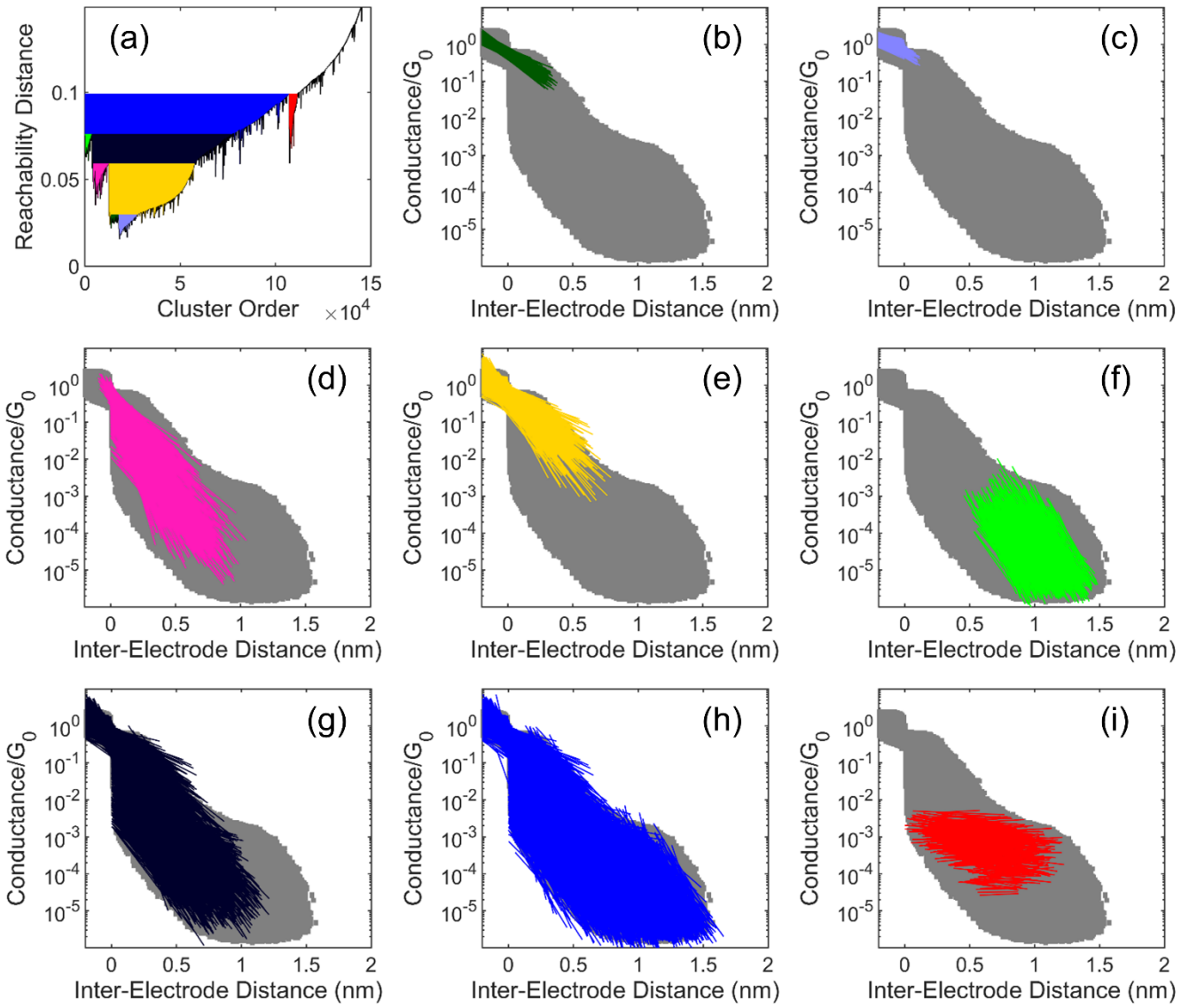

Figure S13. (a) Reachability plot for the OPV2-2BT dataset from Figure 8d with all full-valley clusters hierarchically filled in. (b-i) Segment clusters for each color coded valley from (a), with the cluster in (i) unambiguously identified as the main plateau cluster.

Figure S14 shows full-valley clusters for the C6-2SMe dataset from Figure 8a. The cluster in **Figure S14l** can be unambiguously chosen as the main plateau cluster for this dataset. While the cluster in **Figure S14k** bears a superficial resemblance to the molecular feature, closer inspection reveals that it is much smaller and is composed of very angled segments which are unlikely to correspond to clean molecular

plateaus. A similar main plateau cluster to **Figure S14l** was identified in the other C6-2SMe dataset considered in this work (**Table S3**).

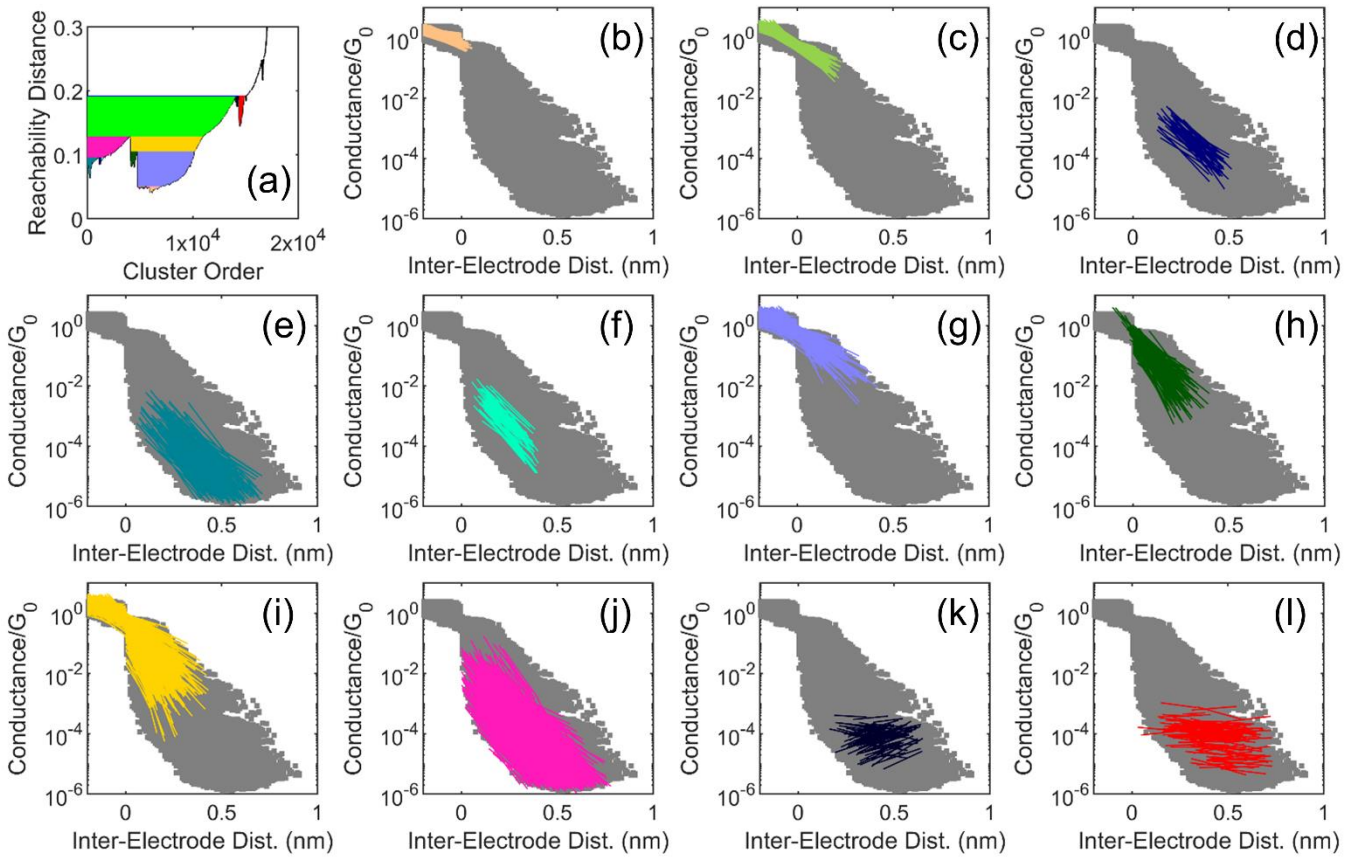

Figure S14. (a) Reachability plot for the C6-2SMe dataset from Figure 8a with all full-valley clusters hierarchically filled in. (b-l) Segment clusters for most of the color coded valleys from (a) (less-important clusters omitted for clarity), with the cluster in (l) identified as the main plateau cluster.

S.11 Cluster Selection for OPV2-2BT/C6-2SMe 1:1 Synthetic Mixture #1

When finding all full-valley clusters for a dataset, the minimum valley size should be set according to the specific context and what types of clusters the user is interested in. For the pure molecular datasets considered in this work, we found that a minimum valley size of 1% of the total number of data points worked well. However, in our synthetic mixture datasets each molecular feature is "diluted" by a factor of two. Moreover, because the C6-2SMe feature is so short, it represents a relatively small number of data points. Therefore, in this context a smaller minimum valley size is appropriate. To demonstrate this,

Figure S15 shows full-valley clusters from the "Mixture #1" dataset from Figure 8g with a minimum valley size of 1%. Although a main plateau cluster can be easily identified (**Figure S15o**), this cluster contains features from both molecules. However, if the minimum valley size is lowered to ~0.5%, then the hierarchical structure produced by Segment Clustering reveals that the cluster from **Figure S15o** is composed of two main sub-valleys (**Figure S15p**). These two sub-valleys represent the clusters shown in Figure 8h, and, as discussed in the main text, correspond to the two different molecular features.

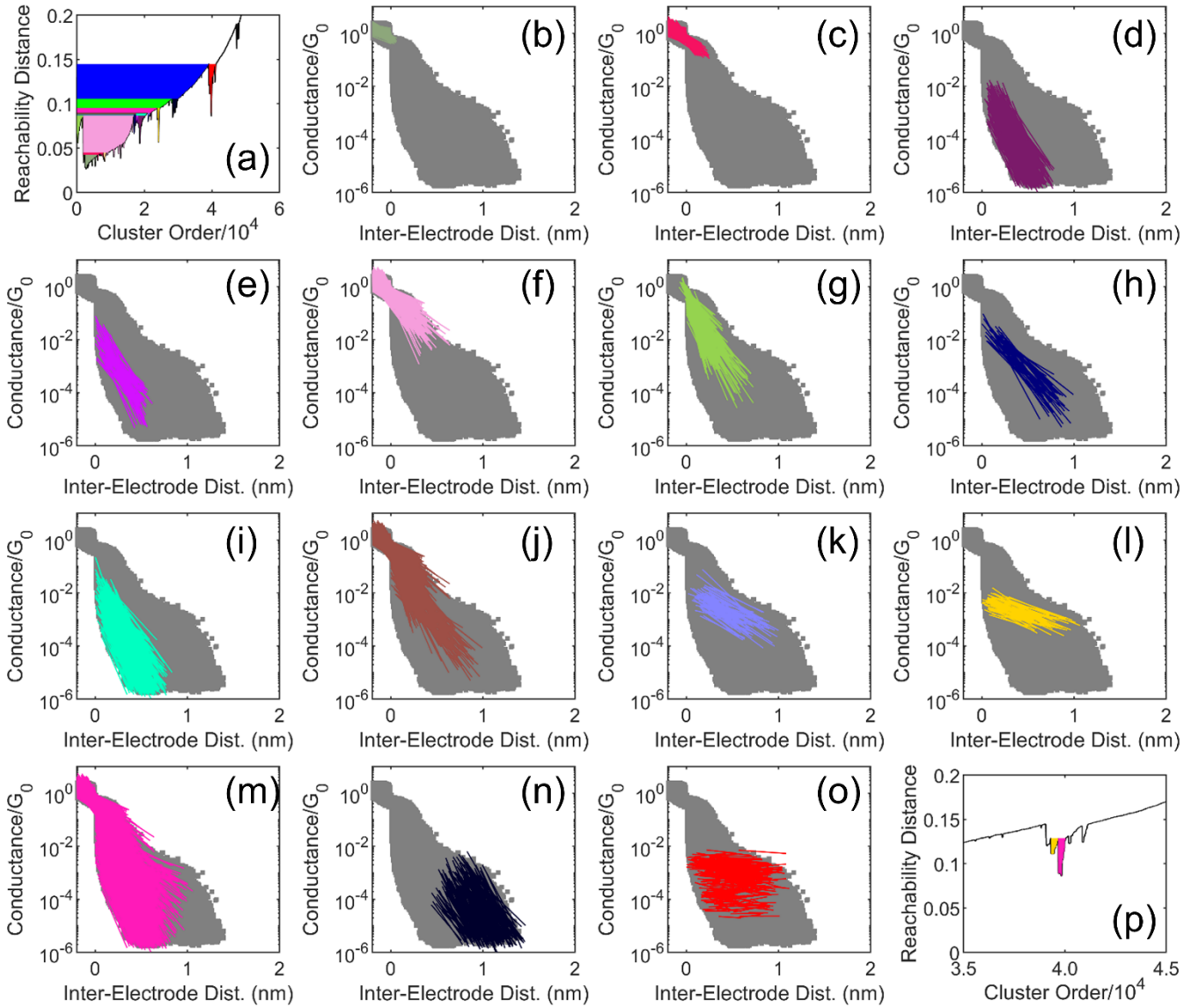

Figure S15. (a) Reachability plot for the Mixture #1 dataset (Fig. 8g) with all full-valley clusters hierarchically filled in. (b-o) Segment clusters for most of the color coded valleys from (a) (less-important

clusters omitted for clarity). The red cluster in (o) is a composite plateau cluster for both molecular features. (p) By lowering the minimum valley size, the cluster in (o) is found to have substructure consisting of two separate valleys, corresponding to the two clusters plotted in Figure 8h.

S.12 Clustering of Additional Synthetic Mixtures

In addition to the OPV2-2BT/C6-2SMe mixture dataset discussed in the main text, 7 additional 1:1 synthetic mixture datasets (for a total of 8) were constructed (see **Table S4** for details) and analyzed in the same way. In seven of these eight total cases, two full-valley clusters were identified that correspond to the main OPV2-2BT and C6-2SMe molecular features (**Figure S16**). Just as with mixture #1 (see section S.11), in each of these cases a "composite" main plateau cluster was first unambiguously identified at the 1% valley size cut-off (analogous to **Figure S15o**), and then lowing of this cut-off revealed two primary sub-valleys (analogous to **Figure S15p**) corresponding to the two molecular features. The clusters identified in this way are shown in **Figure S16**, and their sizes are listed in **Table S6**. The one exception was Mixture #6, where the plateau cluster contained both molecular features did not possess any hierarchical sub-structure (**Figure S16f**). This illustrates the potential drawback of density-based clustering methods mentioned in the main text that dissimilar groups of data may in some cases end up in a single cluster if there is a continuous spread of data between them. We speculate that this issue occurs for this dataset because an error in the attenuation ratios results in similar apparent lengths for both molecules.

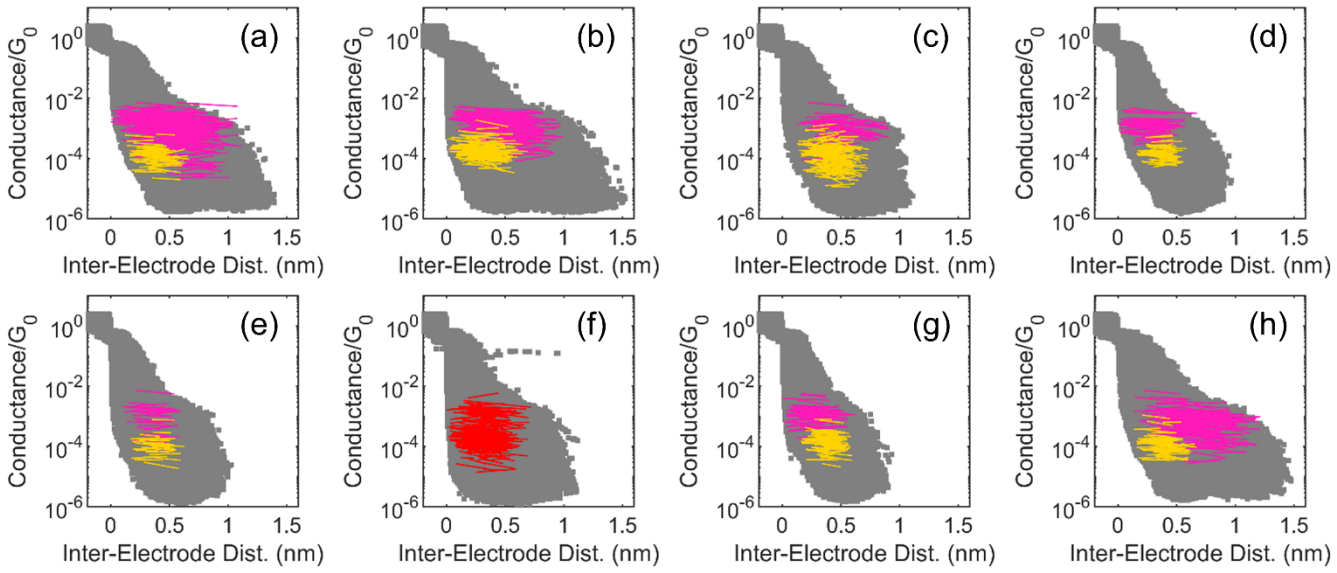

Figure S16. (a-h) Identified molecular plateau clusters for synthetic mixtures #1-8, respectively. In each case, a composite molecular plateau cluster analogous to **Figure S15o** was unambiguously identified (not shown). In 7 out of 8 cases, the valley for that composite cluster was found to contain two sub-valleys, analogous to **Figure S15p**, which were assigned as the OPV2-2BT plateau cluster (pink) and the C6-2SMe plateau cluster (yellow). As shown in **Figure S17** and **Table S6**, these assignments proved to be quite accurate, demonstrating the robustness of Segment Clustering's ability to separate overlapping molecular features. The composite cluster for mixture #6, shown in red in (f), did not contain any hierarchical sub-structure, and so could not be separated.

Just as with Mixture #1 in the main text, each of the OPV2-2BT (C6-2SMe) clusters in **Figure S16** was evaluated for accuracy by calculating how many of the data points assigned to it were from the traces belonging to the OPV2-2BT (C6-2SMe) half of the mixture (**Table S6**). This demonstrates that these separations of overlapping features were successful. While the C6-2SMe clusters again appear to display higher "error rates", as explained in the main text, this is unsurprising given the shorter plateaus for this molecule; the fact that a cluster of short C6-2SMe-like segments is not found in any of the pure OPV2-2BT datasets demonstrates that the source of the "erroneously" included segments is random chance, not mistaken feature identification by the algorithm. Finally, to summarize all of these mixture separation results, **Figure S17** compares the peak conductance values for the two identified molecular clusters from each mixture dataset with the peak conductance values from the main plateau clusters in the pure OPV2-2BT and C6-2SMe datasets.

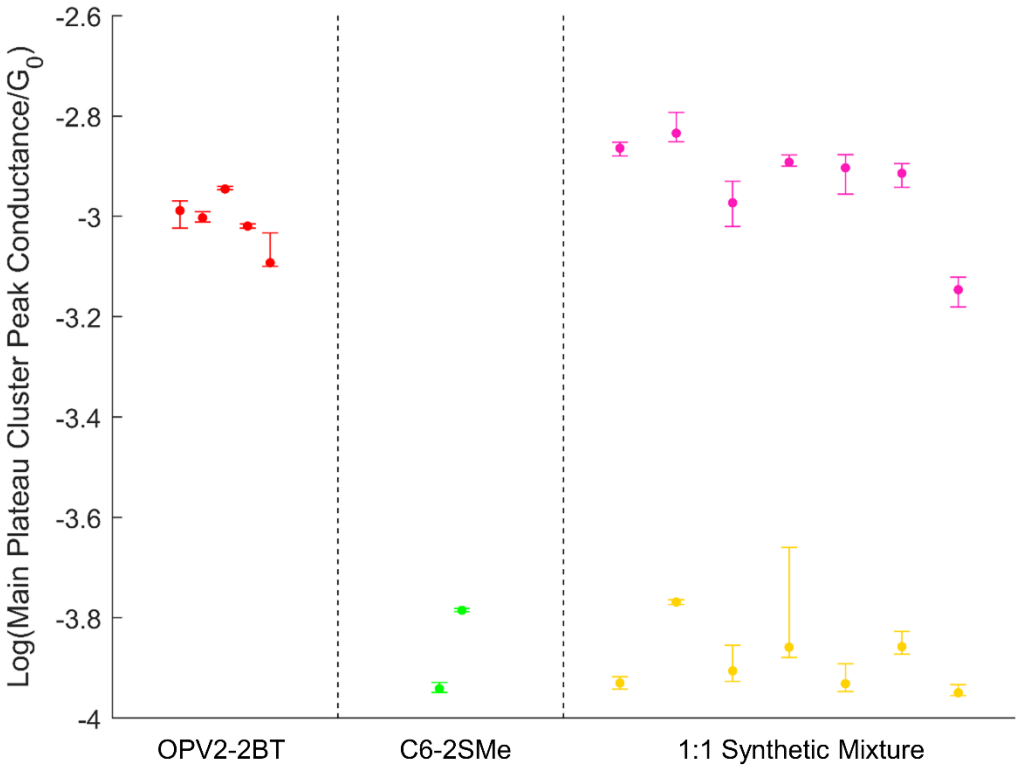

Figure S17. Peak conductance values for the main plateau clusters for the 5 different pure OPV2-2BT datasets considered in this work (red) and the two pure C6-2SMe datasets considered in this work (green). For comparison are plotted the peak conductances of the OPV2-2BT (pink) and C6-SMe (yellow) clusters identified in the seven successfully separated 1:1 synthetic mixture datasets shown in **Figure S16**.

Table S6. For each of the eight mixture datasets considered in this work, the size of the identified C6-2SMe and OPV2-2BT clusters (as a percentage of total data points) and the "accuracy" of each cluster (i.e. how many data points belonging to the cluster come from traces collected in the presence of the molecule that the cluster is assigned to). Each value represents the median from among the twelve different clustering outputs (using different values of the *minPts* parameter) for each dataset. Separate C6-2SMe and OPV2-2BT clusters could not be identified for mixture #6

Mixture #	Data Points Contained in Cluster		Data Points from Correct Half of Dataset	
	C6-2SMe	OPV2-2BT	C6-2SMe	OPV2-2BT
1	0.5%	3.2%	84%	97%
2	1.4%	1.6%	84%	99%
3	1.4%	0.3%	60%	91%
4	0.4%	0.9%	67%	98%
5	1.0%	0.4%	69%	90%
6	NA	NA	NA	NA
7	0.8%	0.7%	59%	96%
8	0.5%	1.8%	76%	95%
AVERAGE	0.9%	1.3%	71%	95%

References

(1) Vrouwe, S. A. G.; van der Giessen, E.; van der Molen, S. J.; Dulic, D.; Trouwborst, M. L.; van Wees, B. J. Mechanics of Lithographically Defined Break Junctions. *Phys. Rev. B* **2005**, *71*, 035313.

(2) Huisman, E. H.; Trouwborst, M. L.; Bakker, F. L.; van Wees, B. J.; van der Molen, S. J. The Mechanical Response of Lithographically Defined Break Junctions. *J. Appl. Phys.* **2011**, *109*, 104305.

(3) Simmons, J. G. Generalized Formula for the Electric Tunnel Effect between Similar Electrodes Separated by a Thin Insulating Film. *J. Appl. Phys.* **1963**, *34*, 1793–1803.

(4) Hong, W.; Manrique, D. Z.; Moreno-García, P.; Gulcur, M.; Mishchenko, A.; Lambert, C. J.; Bryce, M. R.; Wandlowski, T. Single Molecular Conductance of Tolanes: Experimental and Theoretical Study on the Junction Evolution Dependent on the Anchoring Group. *J. Am. Chem. Soc.* **2012**, *134*, 2292–2304.

(5) Grüter, L.; González, M. T.; Huber, R.; Calame, M.; Schönenberger, C. Electrical Conductance of Atomic Contacts in Liquid Environments. *Small* **2005**, *1*, 1067–1070.

(6) Johnson, T. K.; Ivie, J. A.; Jaruvang, J.; Monti, O. L. A. Fast Sensitive Amplifier for Two-Probe Conductance Measurements in Single Molecule Break Junctions. *Rev. Sci. Instrum.* **2017**, *88*, 033904.

(7) Schneider, J.; Vlachos, M. Scalable Density-Based Clustering with Quality Guarantees Using Random Projections. *Data Min. Knowl. Discov.* **2017**, *31*, 972–1005.

(8) Rand, W. M. Objective Criteria for the Evaluation of Clustering Methods. *J. Am. Stat. Assoc.* **1971**, *66*, 846–850.

(9) Salvador, S.; Chan, P. Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms. In *16th IEEE International Conference on Tools with Artificial Intelligence*; 2004; pp 576–584.

(10) Cabosart, D.; El Abbassi, M.; Stefani, D.; Frisenda, R.; Calame, M.; van der Zant, H. S. J.; Perrin, M. L. A Reference-Free Clustering Method for the Analysis of Molecular Break-Junction Measurements. *Appl. Phys. Lett.* **2019**, *114*, 143102.

(11) Hong, W.; Valkenier, H.; Mészáros, G.; Manrique, D. Z.; Mishchenko, A.; Putz, A.; García, P. M.; Lambert, C. J.; Hummelen, J. C.; Wandlowski, T. An MCBJ Case Study: The Influence of π-Conjugation on the Single-Molecule Conductance at a Solid/Liquid Interface. *Beilstein J. Nanotechnol.* **2011**, *2*, 699–713.

(12) Huang, F.; Li, R.; Wang, G.; Zheng, J.; Tang, Y.; Liu, J.; Yang, Y.; Yao, Y.; Shi, J.; Hong, W. Automatic Classification of Single-Molecule Charge Transport Data with an Unsupervised Machine-Learning Algorithm. *Phys. Chem. Chem. Phys.* **2019**, *22*, 1674–1681.

(13) Huang, C.; Jevric, M.; Borges, A.; Olsen, S. T.; Hamill, J. M.; Zheng, J.-T.; Yang, Y.; Rudnev, A.; Baghernejad, M.; Broekmann, P.; et al. Single-Molecule Detection of Dihydroazulene Photo-Thermal Reaction Using Break Junction Technique. *Nat. Commun.* **2017**, *8*, 15436.

(14) Kaliginedi, V.; Moreno-García, P.; Valkenier, H.; Hong, W.; García-Suárez, V. M.; Buiter, P.; Otten, J. L. H.; Hummelen, J. C.; Lambert, C. J.; Wandlowski, T. Correlations between Molecular Structure and Single-Junction Conductance: A Case Study with Oligo(Phenylene-Ethynylene)-Type Wires. *J. Am. Chem. Soc.* **2012**, *134*, 5262–5275.

(15) Moreno-García, P.; Gulcur, M.; Manrique, D. Z.; Pope, T.; Hong, W.; Kaliginedi, V.; Huang, C.; Batsanov, A. S.; Bryce, M. R.; Lambert, C.; et al. Single-Molecule Conductance of

Functionalized Oligoynes: Length Dependence and Junction Evolution. *J. Am. Chem. Soc.* **2013**, *135*, 12228–12240.