Electronic Supplementary Material (ESI) for Journal of Materials Chemistry C.
This journal is © The Royal Society of Chemistry 2023

Supporting information: Generalized neural network approach for separation of molecular breaking traces

Frederik van Veen[1,2], Luca Ornago[1], Herre S.J. van der Zant[1], Maria El Abbassi [1]

[1] *Department of Quantum Nanoscience, Delft University of Technology, 2628CJ Delft, the Netherlands*

[2] *Transport at Nanoscale Interfaces Laboratory, Empa, Swiss Federal Laboratories for Materials Science and Technology, Überlandstrasse 129, CH-8600 Dübendorf, Switzerland*

[*] *email: Frederik.vanVeen@empa.ch, h.s.j.vanderzant@tudelft.nl*

1 Feature vector generation

The feature vectors in this study were constructed from the normalized bin values of the 2D conductance histograms of the individual breaking traces, visualized in figure 1. All traces were initially aligned at the point of junction rupture (chosen to be 0.5 G_0). Subsequently specified ranges of the conductance and electrode displacement axes were divided in logarithmic and linear bins respectively, after which the amount of data points in each bin was counted. Ranges of 0.5-1 $\times 10^{-6}$G$_0$ and 0-2 nm were used, depicted by the green box in figure 1. These ranges were divided into 32 and 25 bins respectively.

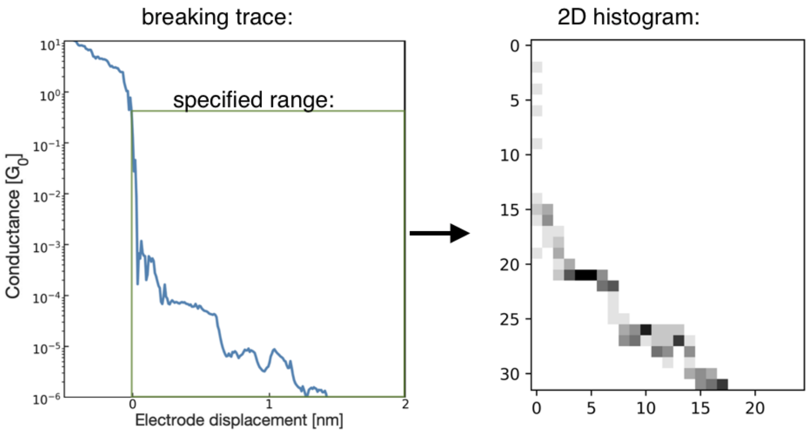

Figure 1: Clustering analysis of the ADT6 traces that were labeled molecular.

2 Full training curve of our CNN

Figure 2 displays the learning curves of the neural network, during training on the full labeled dataset. The learning curves show the accuracy and loss for both the training and validation sets after each complete pass of the training data through the network, while updating its parameters.

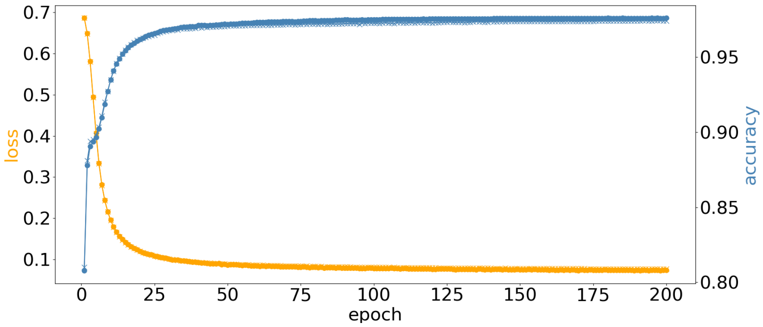

Figure 2: Training curves for our CNN over 200 epochs.

3 Evaluation of the classification performance

In this section, we present the clustering results that were used for a more detailed evaluation of the neural network classification performance.

3.1 ADT6

The classes obtained from a clustering analysis of the molecular and tunnelling sets of ADT6 traces are shown in figures 3 and 4. This clustering is used to assess the performance of the neural network for the separation of tunneling traces presented in the main text figure 3.

3.1.1 Molecular set

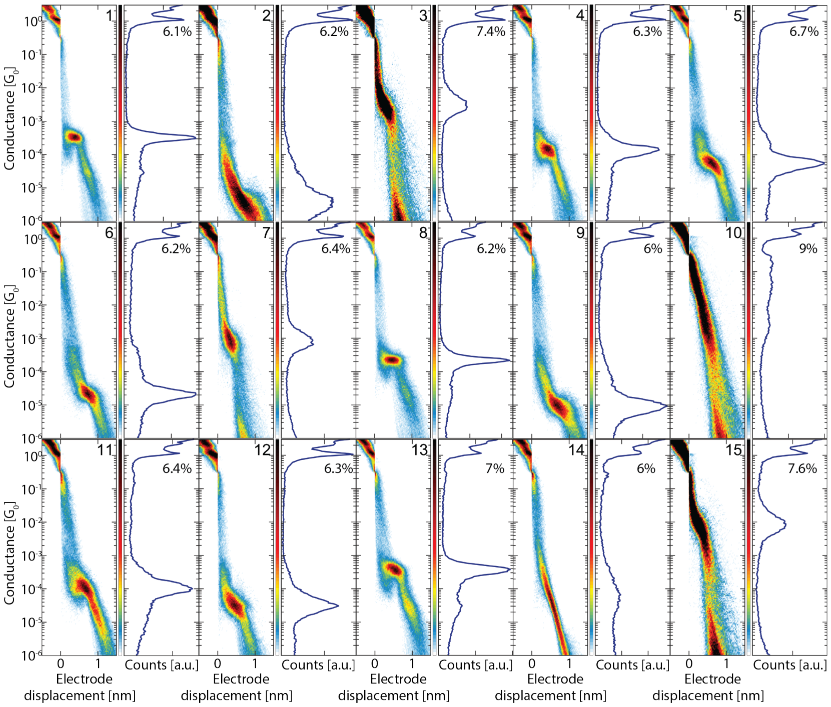

Figure 3: Clustering analysis of the ADT6 traces that were labeled molecular.

3.1.2 Tunnelling set

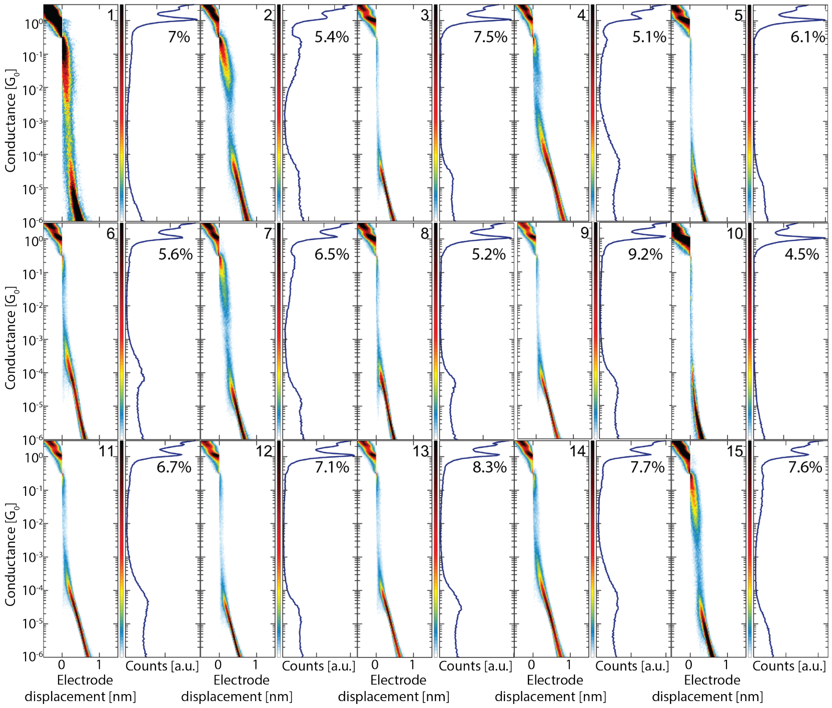

Figure 4: Clustering analysis of the ADT6 traces that were labeled tunnelling.

3.1.3 Estimation of classification accuracy

Of the tunnelling classes, none display any molecular features. Of the molecular classes, class 14 however displays tunneling behaviour. This class contains 1824 traces. The total number of tunneling traces is estimated at 71324, by summing the amount of traces in all tunneling and molecular subclass 14. This shows that 97.5 percent of the tunneling traces are removed while zero molecular ones were discarded. It must be noted that class 10 displays no clean molecular features nor clean tunneling. We have ommited these traces in the calculation of the accuracies.

3.2 OPE3-diSAc

The classes obtained from a clustering analysis of the molecular and tunnelling sets of OPE3-diSAc traces are shown in figures 5 and 6. This clustering is used to assess the performance of the neural network for the separation of tunneling traces presented in the main text figure 3.

3.2.1 Molecular set

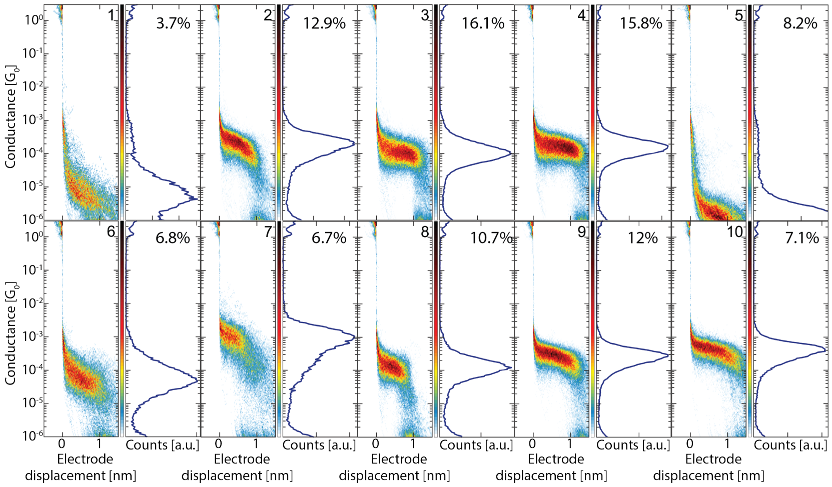

Figure 5: Clustering analysis of the OPE3 traces that were labeled molecular.

3.2.2 Tunnelling set

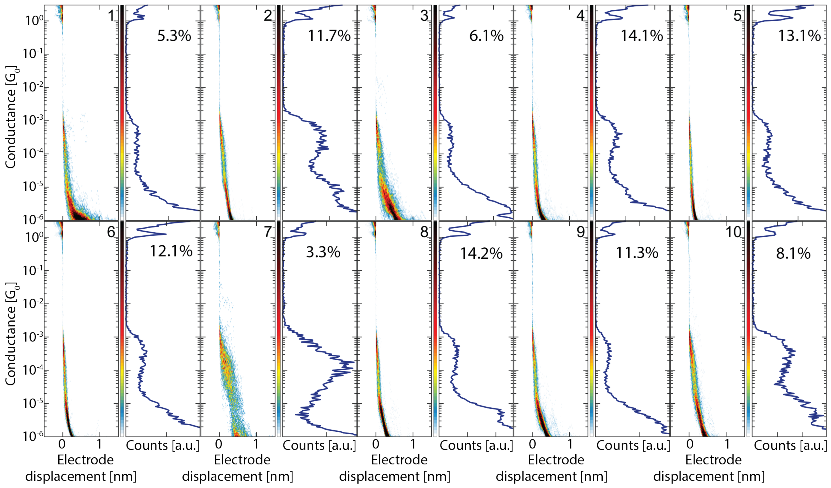

Figure 6: Clustering analysis of the OPE3 traces that were labeled tunnelling.

3.2.3 Estimation of classification accuracy

In the obtained subclasses for the molecular and tunneling sets, it can be seen that none of the traces have been wrongfully labeled by our network. None of the molecular classes display any tunneling traces, while all the tunnelin traces display only clean exponentially decaying traces.

3.3 OPE3-Pyr

The classes obtained from a clustering analysis of the molecular and tunnelling sets of ADT6 traces are shown in figures 7 and 8. This clustering is used to assess the performance of the neural network for the separation of tunneling traces presented in the main text figure 4.

3.3.1 Molecular set

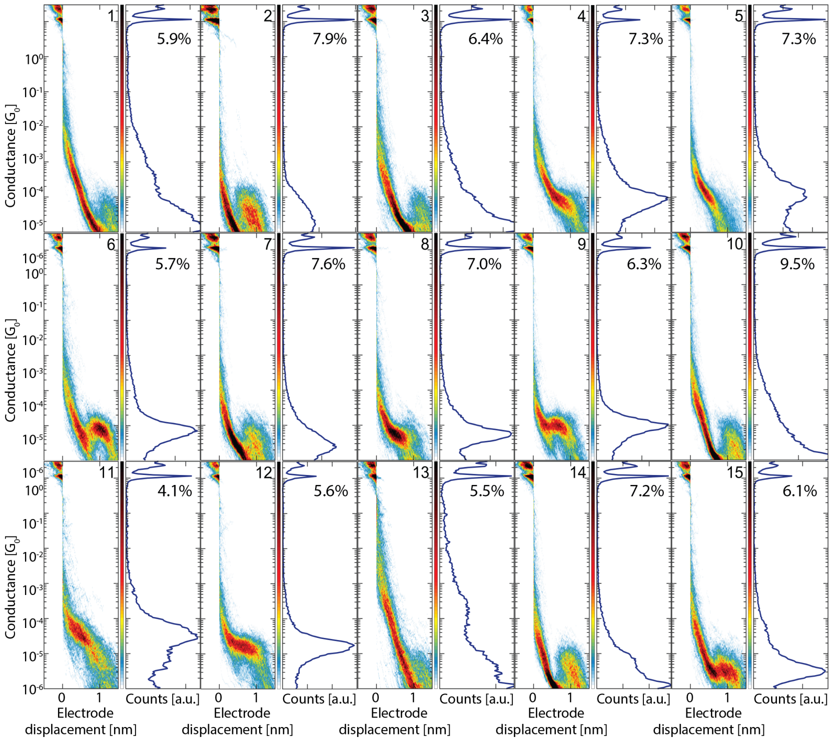

Figure 7: Clustering analysis of the OPE3-Pyr traces that were labeled molecular.

3.3.2 Tunnelling set

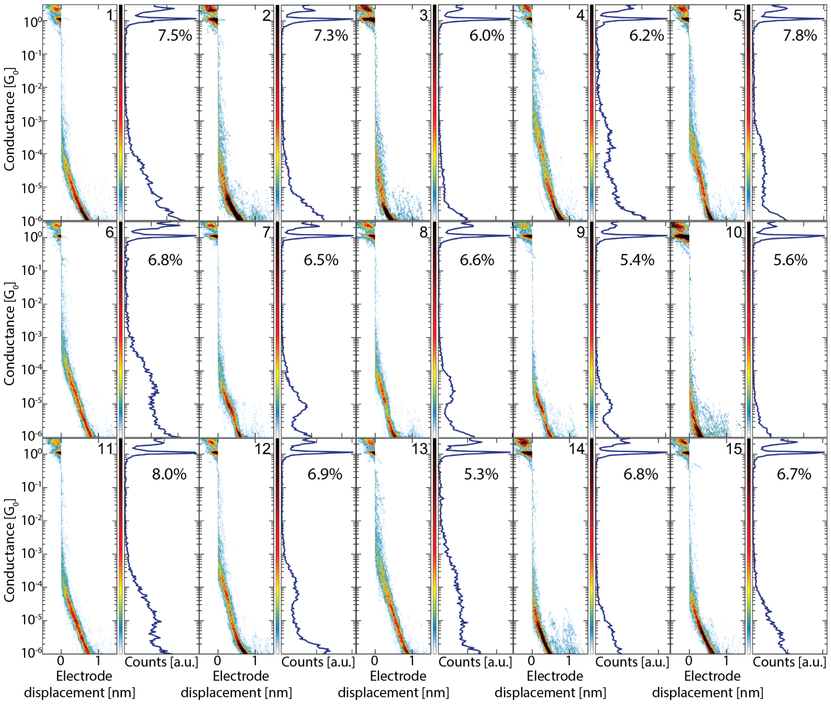

Figure 8: Clustering analysis of the OPE3-Pyr traces that were labeled tunnelling.

3.4 OPE3-NH2

The classes obtained from a clustering analysis of the molecular and tunnelling sets of OPE3-NH2 traces are shown in figures 9 and 10. This clustering is used to assess the performance of the neural network for the separation of tunneling traces presented in the main text figure 4.

3.4.1 Molecular set

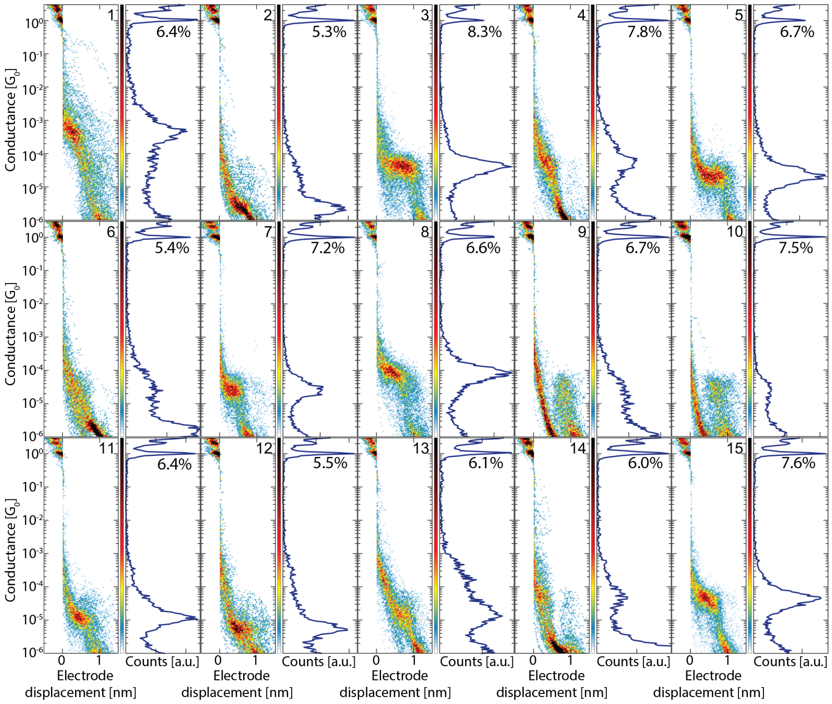

Figure 9: Clustering analysis of the OPE3-NH2 traces that were labeled molecular.

3.4.2 Tunnelling set

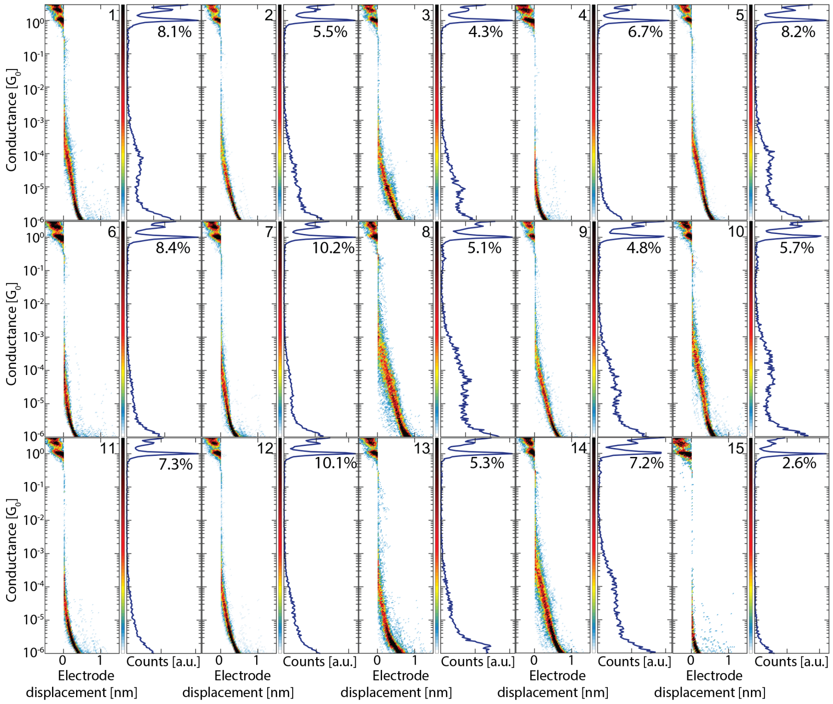

Figure 10: Clustering analysis of the OPE3-NH2 traces that were labeled tunnelling.

4 Benchmarking neural network tunneling-only separation against unsupervised clustering methods

Here we show the full benchmarking results comparing the network performance against commonly used unsupervised techniques (Kmeans and Gaussian mixture model) for the classification of the labeled datasets (ADT3, ADT6, ADT8), with ratios varying from 1:1 to 1:10. The two-class clustering results of this benchmark are shown in Fig. 11 (ADT3), Fig. 12 (ADT6) and Fig. 13 (ADT8).

The green dots are the accuracies obtained with the CNN, while the orange and blue ones were obtained with respectively Gaussian mixtures model and Kmeans clustering.

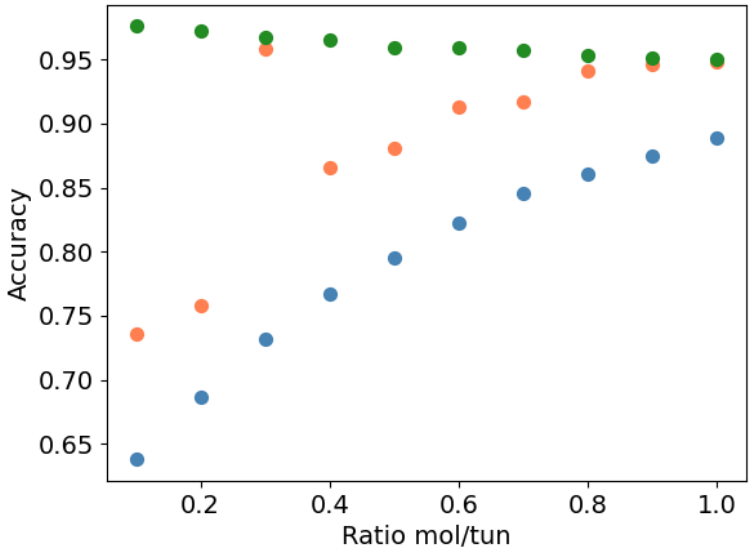

Figure 11: Tunnelling separation performance results of the trained neural network on a labeled ADT3 dataset compared to conventional unsupervised clustering methods.

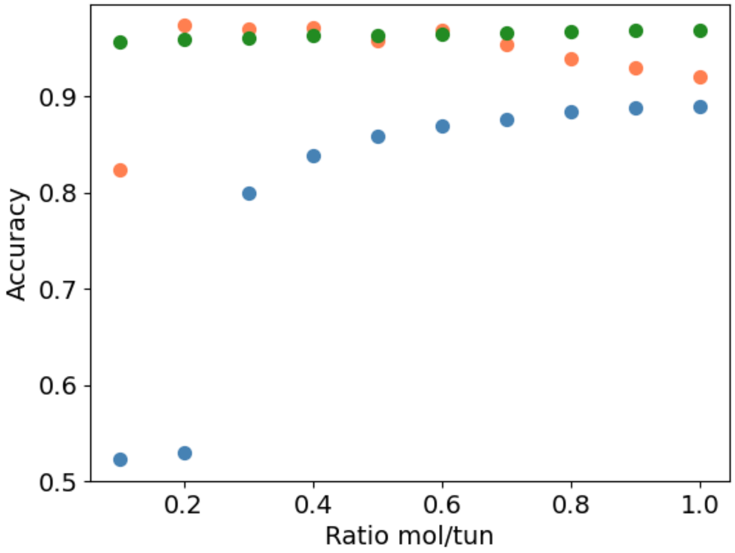

Figure 12: Tunnelling separation performance results of the trained neural network on a labeled ADT6 dataset compared to conventional unsupervised clustering methods.

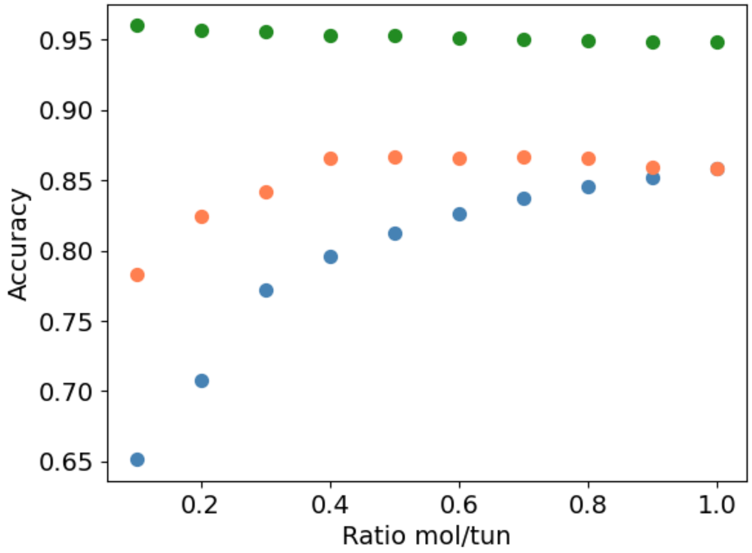

Figure 13: Tunnelling separation performance results of the trained neural network on a labeled ADT8 dataset compared to conventional unsupervised clustering methods.

5 Improvement of Kmeans clustering after separation of tunneling traces

Figure 14 shows the full histograms for the unsupervised clustering of the low-yield ADT3 dataset when tunneling traces are separated (bottom) or not (top). In the main text this figure (Fig. 5) is shown without the 1D histograms for the different classes.

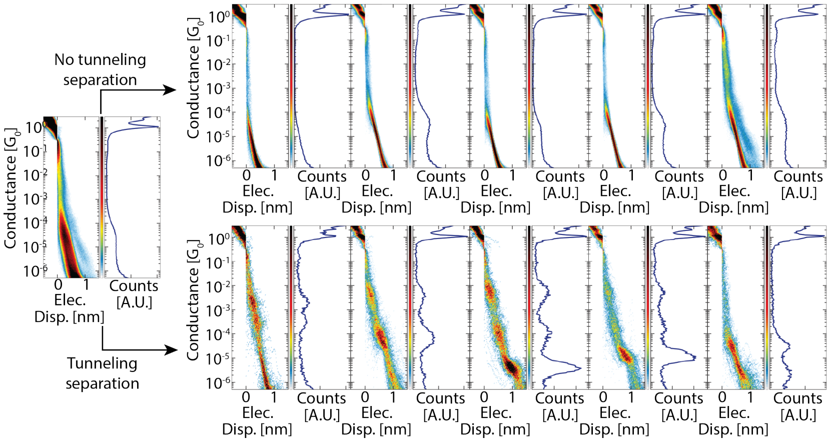

Figure 14: Clustering results on a low molecular-yield dataset (ADT3). Kmeans clustering in five classes without the use of the neural networks to separate the tunneling traces (top) and after removing the tunneling traces by the neural network (bottom).

6 Molecules considered in the study

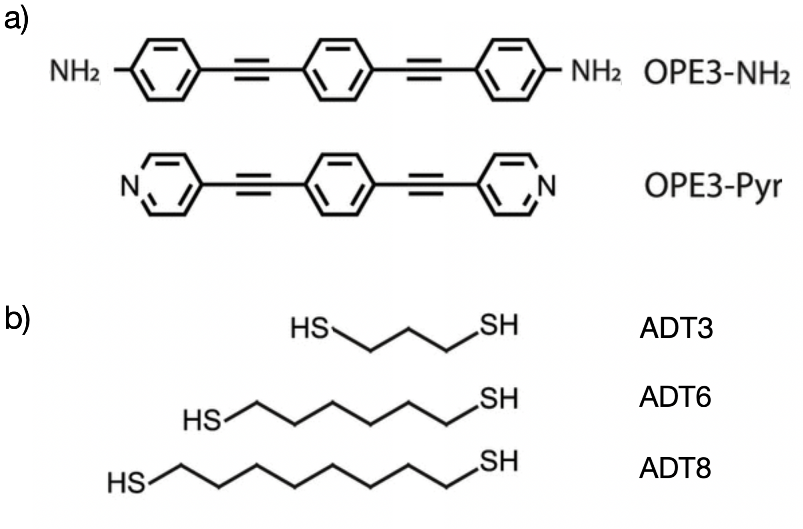

Figure 15: Molecules considered in this this study. OPE3s with different linker groups (a) and alkanedithiols with different chain lengths (b)