Single-Molecule Junction Formation in Break-Junction Measurements

Tianren Fu[1‡], Kathleen Frommer[1‡], Colin Nuckolls[1], Latha Venkataraman*[1,2]

[1] *Department of Chemistry, Columbia University, New York, New York 10027, United States*

[2] *Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York 10027, United States*

Abstract: The scanning tunneling microscope-based break-junction (STM-BJ) technique is the most common method used to study electronic properties of single molecule junctions. It relies on repeatedly forming and rupturing an Au contact in an environment of the target molecules. The probability of junction formation is typically very large (~70-95%) prompting questions relating to how the nanoscale structure of the Au electrode before the metal point-contact ruptures alters junction formation. Here analyze conductance traces measured with the STM-BJ setup by combining correlation analysis and multiple machine-learning tools, including gradient boosted trees and neural networks. We show that two key features describing the Au-Au contact prior to rupture determine the extent of the contact relaxation (the snapback) and the probability of junction formation. Importantly, our data indicates strongly that molecular junctions are formed prior to the rupture of the Au-Au contact, explaining the high probability of junction formation observed in room-temperature solution measurements.

The scanning tunneling microscope-based break-junction (STM-BJ) technique has proven to be a unique and versatile tool for investigating the physio-chemical properties of single metal-molecule-metal junctions.[1-2] STM-BJ technique can robustly construct and characterize single molecular junctions of with molecules ranging from organic, inorganic and bio-molecules.[3-8] It is also versatile in that it can be used to measure electronic, mechanical, thermoelectric and photoconducting properties of the junctions.[9-14] In STM-BJ experiments, the impact that the nanoscale electrode structure and its evolution and relaxation upon elongation and rupture play on the molecular junction formation is not well studied or well understood.[15-17] Recently, machine learning-assisted analyses have demonstrated the ability to analyze break-junction data to gain insights into molecular junction properties.[18-22] Here we employ machine learning techniques, from simple correlation analysis to deep neural networks, to comprehensively analyze this problem, and show that we can learn more about the underlying factors that make STM-BJ measurement method robust, reliable and reproducible.

In the STM-BJ method, metal-molecule-metal junctions are repeatedly formed and elongated until they break, while the current across the junction is continuously measured under an applied bias voltage, producing a conductance versus distance trace. At the start of such a trace, the metal electrodes are in contact, resulting in a high conductance, and as the STM tip is pulled away, the conductance drops in steps until a value close to 1 G_0, where $G_0 = 2e^2/h$ is the conductance quantum. This indicates the formation of a single atomic Au-Au contact, which breaks upon further elongation. Following its rupture, a single-molecule junction conductance plateau is often observed indicating that a molecule bridges the gap between the electrodes.[1-2] The average molecular plateau length is related to the molecular backbone length, however the plateau length varies significantly from trace to trace and can depend on the molecular configuration within the junction[15]. It could also be related to the junction formation probability which can depend on the linker groups[23-24]. This average plateau length, however, is not equal to the length of the molecular junction. The difference is often attributed to the fact that Au electrodes reorganize and relax when point-contact ruptures opening up a gap, known as the "snapback" distance, and this is used to account for the difference between the molecular junction length and the plateau length.[25] Usually, this snapback distance is reported as a single value[23, 25-27]. Here, we show that the snapback distance is affected by the structure of the Au contact formed prior to the formation of the molecular junction which is altered by the solvent, and thus depends on the experimental

conditions. Importantly, we also show that for individual traces, the measured snapback is not strongly correlated with the plateau length of a molecular junction, indicating that the plateau length is much less sensitive to the contact formation history.

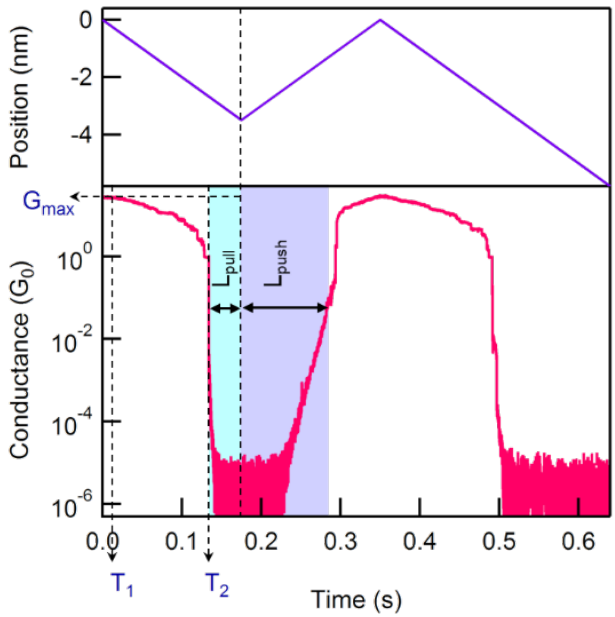

Figure 1. Sample piezo ramp (upper panel) and conductance versus time trace (lower panel) for a push-pull trace measured in 1,2,4-trichlorobenzene (TCB) at an applied bias of 100 mV and pulling rate of 15 nm/s. T_1 indicates the time when the highest conductance (G_{max}) is observed; T_2 is the time when the initial Au contact ruptures. L_{pull} indicates the distance pulled after the contact breaks and L_{push} indicates the distance pushed before the contact is reformed.

To probe the structure of the Au contact and determine snapback distances, we modified the standard STM-BJ measurement. The Au contact is initially pulled apart, then pushed back to remake contact, and finally pulled apart again. A sample "pull-push"[26, 28] conductance trace with the accompanying voltage ramp applied to the piezoelectric transducer that controls the substrate motion relative to the tip is shown in Figure 1 plotted against time. These measurements are made in a solvent, 1,2,4-trichlorobenzene (TCB), on a Au-coated substrate. As indicated in Figure 1, the time at which the Au-Au contact has the highest conductance (G_{max}) is designated as T_1 and occurs before a single atomic contact forms. This single-atom Au-Au contact breaks at time T_2, and this corresponds to the time when a large conductance drop is seen just below 1 G_0. The displacement at T_2 is denoted as L_{break}. Beyond T_2, conductance drops to the instrument noise floor (~ 10^{-5} G_0) if there is no molecule and remains at this level through the end of the pulling phase (light blue

shaded region in Figure 1 designated L_{pull}) and into the beginning of the pushing phase of the measurement. On pushing the electrodes further together (dark blue shaded region in Figure 1 designated L_{push}), a contact is formed again. We use a conductance threshold of 0.05 G_0 to indicate contact formation (although other thresholds up to 1 G_0 do not alter our findings). We see that the time required to reform a Au-Au contact while pushing is greater than the time the electrodes are pulled apart after breaking the contact (i.e. $L_{push} > L_{pull}$). This is due to the snapback reflecting that the Au electrodes reorganize and relax after the contact is broken. The snapback is defined as $L_{push} - L_{pull}$. For our analysis, we also consider three additional features related to the Au-Au contact evolution: the slope of the conductance versus distance trace for the Au region (m_{Au}), which is determined by doing a linear regression on the region between T_1 and T_2; the length of the plateau around 1 G_0 (L_1) and length of the plateau around 2 G_0 (L_2). If the measurements are done in a solution of molecules, after T_2, instead of dropping into noise floor, a molecular conductance plateau is observed (see SI Figure S1 for a sample trace). The length of this plateau in a trace is the distance between the first and last point in the trace that is within the molecular conductance region as determined from a one-dimensional conductance histogram. These five parameters, G_{max}, L_1, L_2, L_{break} and m_{Au} describe evolution of the Au contact which we use to analysis the relation between snapback, molecular plateau length and the Au contact formation history.

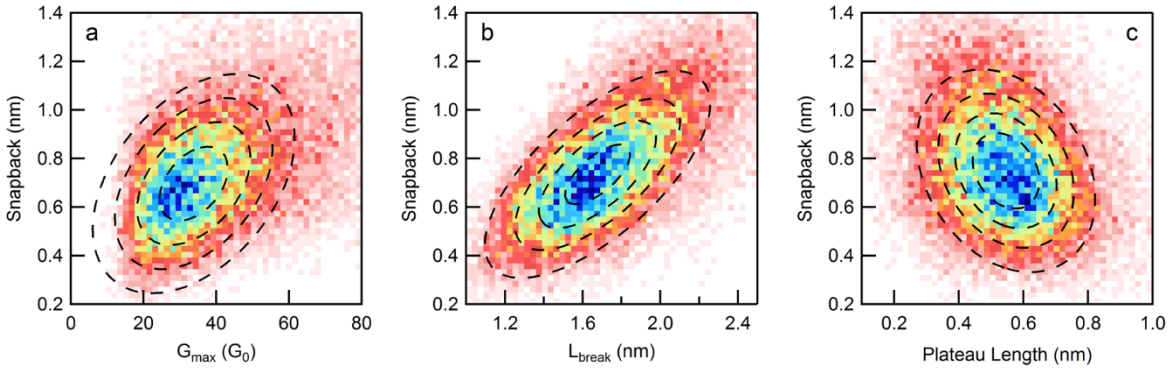

Figure 2. Two-dimensional (2D) correlation histograms constructed from 24,880 selected push-pull traces of 4,4"-diamino-*p*-terphenyl 0.1 mM in TCB solution. Black dashed lines are contour lines of 2D Gaussian fits. Snapback versus **(a)** G_{max} with a Pearson correlation coefficient of 0.412; **(b)** L_{break} with a Pearson correlation coefficient is 0.679 and **(c)** Molecular plateau length with a Pearson correlation coefficient is -0.265. See SI Figure S4e-h for histograms of raw data.

To demonstrate the correlation between these five parameters, two-dimensional correlation histograms are constructed from 24,880 measurements with 4,4"-diamino-*p*-terphenyl from a TCB solution. Figures 2a and 2b shows the correlation between snapback and G_{max} and L_{break}, respectively. We see that snapback is positively correlated with both parameters though the correlation between snapback and L_{break} is much stronger. To rationalize this finding, we note that in addition to structural changes, the force required to elongate a contact also stretches the Au-Au bonds within the tip asperities and the single atomic contact.[15] Like a spring, which will recoil more the more it is stretched, elongating the bonds over a larger distance (a larger L_{break}) will result in a larger relaxation upon rupture, resulting in a larger snapback. Moreover, a larger G_{max} indicates that the contact cross-section has many more atoms, and such a thicker contact will require a longer elongation to break resulting in a larger snapback. Figure 2c shows the correlation plot between snapback and molecular plateau length and reveals they are very weakly and negatively correlated. Since a larger snapback results in a wider gap right after the rupture of Au-Au point contact, it is reasonable that larger snapback reduces the further displacement needed to break the molecular junction, and results in a shorter plateau length. The small magnitude of correlation, however, is surprising. It indicates that the Au contacts relax fully only after the rupture of Au-molecule-Au junction. This can be rationalized by considering that the molecule can provide a force necessary to hold the electrodes in a slightly stretched form preventing them from relaxing as illustrated in SI Figure S5. However, this requires the molecule to be bridging across the electrodes even before the Au-Au point contact breaks, otherwise the relaxation will occur before the molecular bridge forms. This picture is indeed consistent with molecular dynamic simulations.[29-30]

To fully understand the impact of the Au-Au contact evolution history on snapback and plateau length, we plot in Figure 3a, the absolute value of their correlations with the five measured parameters: we can see that snapback depends primarily on G_{max} and L_{break}, and not strongly on L_1, L_2 and m_{Au}. However, none of these parameters are strongly correlated with molecular junction plateau length, indicating the plateau length is not determined by the evolution of the Au contact prior to rupture.

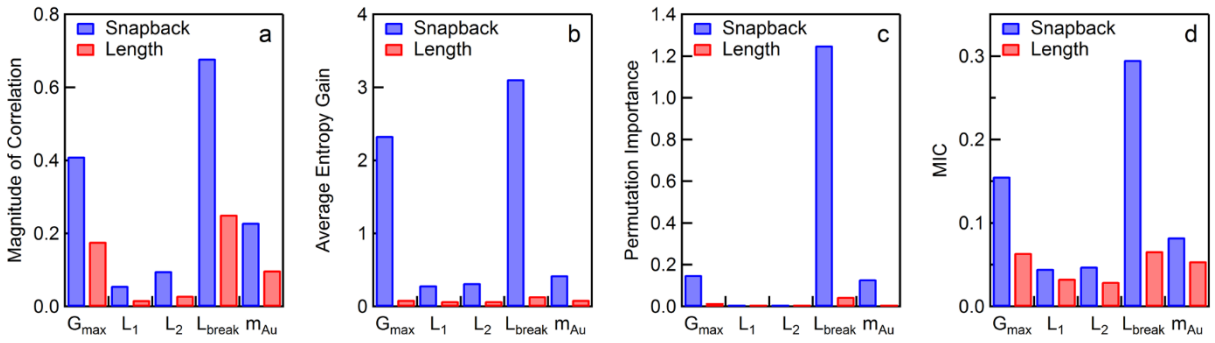

Figure 3. Metrics characterizing the importance of different parameters in determining snapback (blue bars) and molecular plateau length (red bars), for measurements in a 0.1 mM solution of 4,4"-diamino-*p*-terphenyl in TCB. The metrics are: (a) the magnitude of the correlation, (b) the total information entropy gain during the training of XGBoost models, (c) the permutation importance according to XGBoost models and (d) the mutual information coefficient (MIC).

The correlation analysis, however, only interprets the linear relations between these parameters. In order to confirm these findings and to see if there might be some nonlinear relations that could change the conclusion, we also apply a few other methods to characterize the importance of these five parameters on snapback and the molecular junction plateau length. The first method, gradient boosted trees (GBT)[31-32] is a machine learning algorithm with high expressivity and generalizability. GBT can find non-linear relations between parameters and has been widely used for feature extraction and selection.[33-35] For the analysis here, we use the XGBoost[36] package, which is one of the most powerful and frequently used implementation of GBT. In a typical GBT model, many decision trees are constructed to determine the dependent variable (say snapback or plateau length) from the independent variables (G_{max}, L_{break}, L_1, L_2 and m_{Au}). Each decision tree is made of many if-then-else decision nodes and the path taken through these nodes determines the output of a tree. During the training process, these nodes are built recursively on the independent variables (a process known as splitting) to satisfy the maximization of information entropy gain after applying the corresponding if-then-else rules.

We show in Figure 3b the importance of each parameter in determining snapback or plateau length. This importance is evaluated as the average information entropy gain for a parameter over all the splitting done during the tree construction process. We see that G_{max} and L_{break} are important parameters in predicting snapback and no parameter that describes the gold contact structure predicts the molecular plateau length. With this XGBoost model, we also measure the permutation importance of each feature, shown in Figure 3c. The permutation importance[37-38] is a robust metric

against bias on a parameter distribution or model design; permutation importance of one parameter is defined as the performance drop of the model when we randomly shuffle the parameter to make it irrelevant. If the parameter is important, the model will perform worse without it. Here, we find that L_{break} is very helpful in determining the snapback while all other parameters are not critical.

In Figure 3d, we show the maximum information coefficients (MIC)[39-41] between the parameters and snapback or plateau length using the *minepy* package.[39] MIC measures the dependence between two parameters that are either linearly or non-linearly related; it reflects the noise level in the data regardless of what the actual underlying relation between the parameters is. Again, we can see that L_{break} has a high importance in determining the snapback, G_{max} has some importance, but the other three are negligibly important. For the plateau length, however, none of the parameters are important. This confirms our earlier hypothesis that molecular junctions are formed prior to the rupture of the Au-Au contact.

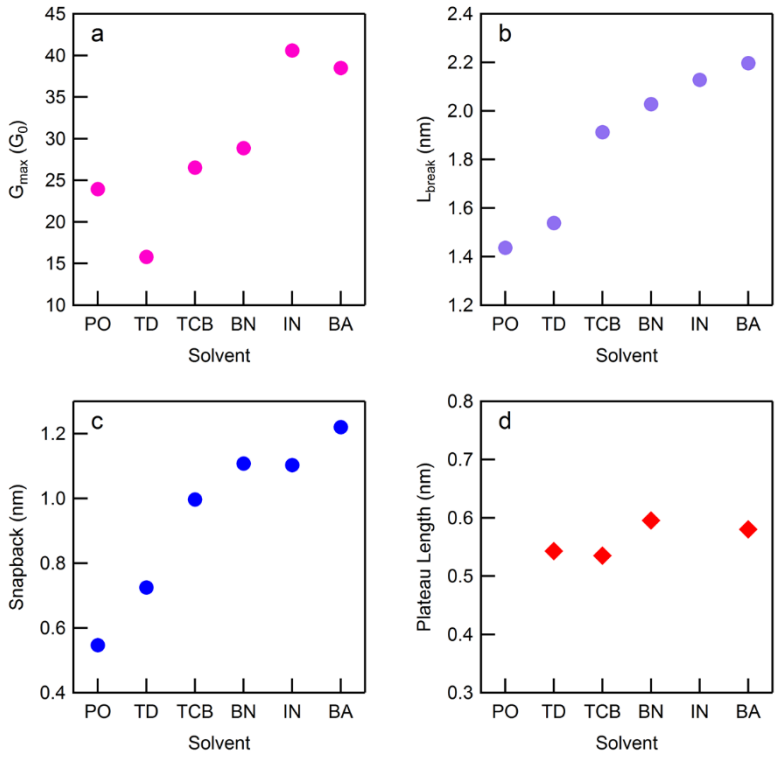

Figure 4. The most probable **(a)** G_{max}, **(b)** L_{break}, and **(c)** snapback for 6 different solvents determined from Gaussian fits to histogram data, from pure solvent measurements; **(d)** The molecular plateau length for 4,4"-diamino-*p*-terphenyl solutions of 4 different solvents. Abbreviations for solvents used are as follows: PO = 1-phenyloctane, TD = tetradecane, TCB = 1,2,4-trichlorobenzene, BN = 1-bromonaphthalene, IN = 1-iodonaphthalene, BA = 4-bromoanisole. See SI Figure S2 and S4 for raw data.

To determine if our results are limited by the fact that we use just five manually-selected parameters, we try next to see if we can predict the snapback and plateau lengths by exploiting all the information on the Au contact evolution history, i.e. the trace through T_2. For this, we build a convolutional neural network (CNN)-based deep learning model. Deep learning algorithms on conductance traces have been proven effective and to be able to extract information other than basic features like length and mean conductance.[42-44] By constructing two models with the identical structure to predict the snapback and plateau length, we find that the correlation between prediction and the actual value is 73.1% for snapback, and only 32.4% for plateau length (see SI section 3 for details). This indicates that CNN algorithm also recognizes the weak correlation between the molecular junction plateau length and the Au contact evolution prior to junction formation.

We now turn to measurements made in different solvents to understand how the environment affects the elongation and rupture of Au contacts. Figure 4a shows G_{max} determined from a series of measurements in different solvents commonly used for STM-BJ measurements, all obtained from commercial sources (see Figure 4 caption for list). Figure 4b shows L_{break} for these same solvents and Figure 4c shows the measured snapback. From Figures 4a-c we see that G_{max}, L_{break} and snapback follow the same trends. We find that solvents with a low snapback value, such as phenyloctane (PO) and tetradecane (TD), are those which interact weakly with Au electrodes, and solely through Van der Waals interactions. By contrast, solvents with high snapback values, such as 1-bromonaphthalene (BN), 1-iodonaphthalene (IN) and 4-bromoanisole (BA), interact more strongly with the soft Au atoms through their soft halide group. Solvent-Au binding energy calculations[45] confirm this finding. Since these solvent molecules bind to undercoordinated Au atoms that are pulled out of the surface, they stabilize the newly-formed Au surface, and thus decrease the energy required to elongate the Au contact. This in turn allows for a longer L_{break}. Since we have shown above that L_{break} is positively correlated with the snapback distance, solvents that passivate undercoordinated Au atoms are likely to lead to a longer snapback. In Figure 4d we show the measured molecular plateau length of 4,4"-diamino-*p*-terphenyl solution in TD, TCB, BN and BA. We find the plateau lengths are similar across different solvents and do not follow the trend seen with the snapback, likely because the linker-Au interaction is much less affected by the solvent effect. Additionally, this is consistent with our finding that the plateau length is very weakly correlated to the G_{max}, L_{break} and snapback.

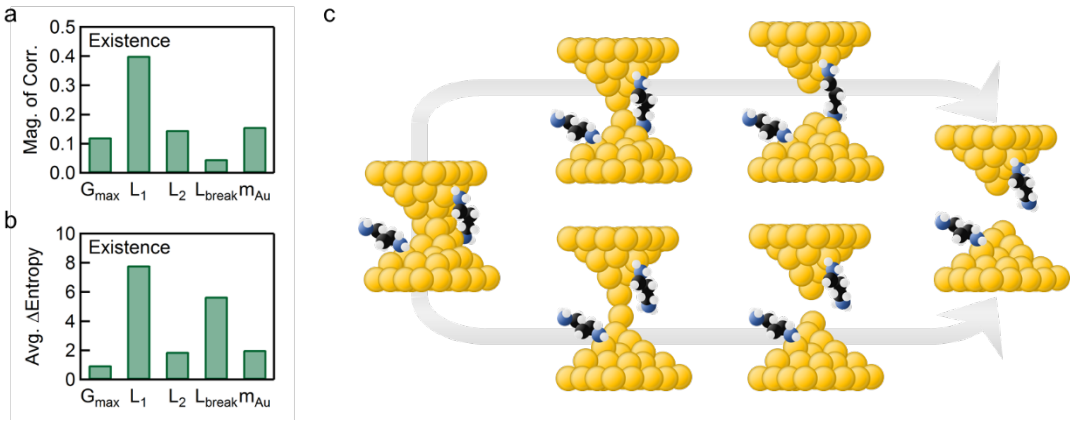

Figure 5. (a) The importance of different features in determining the existence of molecular junction (green bars) for the TCB solution of 1,3-propanediamine, as the magnitude of correlation (upper) and the tree-splitting importance according of the XGBoost models (lower). **(b)** Illustration of the pathway for the rupture of a short (upper) and long (lower) $1G_0$ contact with a short molecule in parallel to the contact. A molecular junction remains only in the upper path with a short Au-Au contact.

For long molecules like 4,4''-diamino-*p*-terphenyl, nearly every Au-contact that is ruptures forms a molecular junction, so we turn to 1,3-diaminopropane to see if the junction formation probability differs in shorter molecules. We repeat the modified break-junction measurement illustrated in Figure 1 in a solution of 1,3-diaminopropane in TCB and analyze our data. Figure 5a and 5b shows the correlation (absolute values) and tree-splitting importance (average information entropy gain of the XGBoost model) of junction formation probability versus G_{max}, L_{break}, L_1, L_2 and B_{Au}. We can see that the junction formation probability is negatively correlated with the L_1 (the length of the $1G_0$ plateau), but is much less correlated with any other parameter. This indicates that junctions with smaller L_1 have a higher chance of forming a molecular junction while those with a longer L_1 are less likely to form a molecular junction. This is consistent with our earlier hypothesis that the molecules bind to the two electrodes in parallel to the gold point-contact. For junctions with short L_1, a pre-existing molecular bridge is likely to survive after the Au point-contact ruptures as illustrated in the upper pathway in Figure 5c. For junctions with longer L_1, the molecular bridge is likely to rupture before the Au-Au point contacts ruptures as illustrated in the lower pathway in Figure 5b. Together, these findings confirm our hypothesis that molecular junctions form prior to the rupture of the metal-contact in STM-BJ measurements.

In conclusion, through modified STM-BJ experiments we have shown that the relaxation of Au electrodes after breaking a point contact depends on the environment around the gold

electrode. We show that this snapback is mainly dependent on two parameters that describe the Au contact prior to rupture: G_{max} (highest conductance) and L_{break} (displacement until rupture). We however find that the molecular plateau length is only weakly and negatively correlated to the snapback, and is nearly independent on the parameters describing the Au contact. We find that the molecular junction plateau length and the junction formation probability for short molecules is mostly independent on the Au contact structure prior to rupture but we do find that it is negatively correlated with the length of the 1-G_0 plateau. These results indicate that the molecules are bound to the Au electrode before the Au point contact ruptures. A complete relaxation of the electrodes happens only after the molecular junction also ruptures. These findings provide key insights into the versatility of STM-BJ measurements to form and characterize molecular conductance signatures in a ranger of solvents and environments.

ASSOCIATED CONTENT

Supporting Information

Additional figures, synthetic details and model details. The Supporting Information is available free of charge on the ACS Publications website.

AUTHOR INFORMATION

Corresponding Author

lv2117@columbia.edu

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

We acknowledge financial support from the National Science Foundation under award grant CHE-1764256 and DMR-1807580.

References:

1. Venkataraman, L.; Klare, J. E.; Tam, I. W.; Nuckolls, C.; Hybertsen, M. S.; Steigerwald, M. L., Single-molecule circuits with well-defined molecular conductance. *Nano Lett* **2006,** *6* (3), 458-62.
2. Xu, B.; Tao, N. J., Measurement of single-molecule resistance by repeated formation of molecular junctions. *Science* **2003,** *301* (5637), 1221-3.

3. Brisendine, J. M.; Refaely-Abramson, S.; Liu, Z. F.; Cui, J.; Ng, F.; Neaton, J. B.; Koder, R. L.; Venkataraman, L., Probing Charge Transport through Peptide Bonds. *J Phys Chem Lett* **2018,** *9* (4), 763-767.
4. Milan, D. C.; Krempe, M.; Ismael, A. K.; Movsisyan, L. D.; Franz, M.; Grace, I.; Brooke, R. J.; Schwarzacher, W.; Higgins, S. J.; Anderson, H. L.; Lambert, C. J.; Tykwinski, R. R.; Nichols, R. J., The single-molecule electrical conductance of a rotaxane-hexayne supramolecular assembly. *Nanoscale* **2017,** *9* (1), 355-361.
5. Wang, L.; Gong, Z. L.; Li, S. Y.; Hong, W.; Zhong, Y. W.; Wang, D.; Wan, L. J., Molecular Conductance through a Quadruple-Hydrogen-Bond-Bridged Supramolecular Junction. *Angew Chem Int Ed Engl* **2016,** *55* (40), 12393-7.
6. Su, T. A.; Neupane, M.; Steigerwald, M. L.; Venkataraman, L.; Nuckolls, C., Chemical principles of single-molecule electronics. *Nature Reviews Materials* **2016,** *1* (3).
7. Xiang, L.; Palma, J. L.; Bruot, C.; Mujica, V.; Ratner, M. A.; Tao, N., Intermediate tunnelling-hopping regime in DNA charge transport. *Nat Chem* **2015,** *7* (3), 221-6.
8. Zotti, L. A.; Bednarz, B.; Hurtado-Gallego, J.; Cabosart, D.; Rubio-Bollinger, G.; Agrait, N.; van der Zant, H. S. J., Can One Define the Conductance of Amino Acids? *Biomolecules* **2019,** *9* (10).
9. Aradhya, S. V.; Venkataraman, L., Single-molecule junctions beyond electronic transport. *Nat Nanotechnol* **2013,** *8* (6), 399-410.
10. Huang, C.; Rudnev, A. V.; Hong, W.; Wandlowski, T., Break junction under electrochemical gating: testbed for single-molecule electronics. *Chem Soc Rev* **2015,** *44* (4), 889-901.
11. Rincon-Garcia, L.; Evangeli, C.; Rubio-Bollinger, G.; Agrait, N., Thermopower measurements in molecular junctions. *Chem Soc Rev* **2016,** *45* (15), 4285-306.
12. Gehring, P.; Thijssen, J. M.; van der Zant, H. S. J., Single-molecule quantum-transport phenomena in break junctions. *Nature Reviews Physics* **2019,** *1* (6), 381-396.
13. Moreno-Garcia, P.; Gulcur, M.; Manrique, D. Z.; Pope, T.; Hong, W.; Kaliginedi, V.; Huang, C.; Batsanov, A. S.; Bryce, M. R.; Lambert, C.; Wandlowski, T., Single-molecule conductance of functionalized oligoynes: length dependence and junction evolution. *J Am Chem Soc* **2013,** *135* (33), 12228-40.
14. Reddy, P.; Jang, S. Y.; Segalman, R. A.; Majumdar, A., Thermoelectricity in molecular junctions. *Science* **2007,** *315* (5818), 1568-71.
15. Kamenetska, M.; Koentopp, M.; Whalley, A. C.; Park, Y. S.; Steigerwald, M. L.; Nuckolls, C.; Hybertsen, M. S.; Venkataraman, L., Formation and evolution of single-molecule junctions. *Phys Rev Lett* **2009,** *102* (12), 126803.
16. Quek, S. Y.; Venkataraman, L.; Choi, H. J.; Louie, S. G.; Hybertsen, M. S.; Neaton, J. B., Amine-gold linked single-molecule circuits: experiment and theory. *Nano Lett* **2007,** *7* (11), 3477-82.
17. Huisman, E. H.; Trouwborst, M. L.; Bakker, F. L.; de Boer, B.; van Wees, B. J.; van der Molen, S. J., Stabilizing single atom contacts by molecular bridge formation. *Nano Lett* **2008,** *8* (10), 3381-5.
18. Bamberger, N. D.; Ivie, J. A.; Parida, K. N.; McGrath, D. V.; Monti, O. L. A., Unsupervised Segmentation-Based Machine Learning as an Advanced Analysis Tool for Single Molecule Break Junction Data. *The Journal of Physical Chemistry C* **2020,** *124* (33), 18302-18315.

19. 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. *Applied Physics Letters* **2019,** *114* (14).
20. Hamill, J. M.; Zhao, X. T.; Meszaros, G.; Bryce, M. R.; Arenz, M., Fast Data Sorting with Modified Principal Component Analysis to Distinguish Unique Single Molecular Break Junction Trajectories. *Phys Rev Lett* **2018,** *120* (1), 016601.
21. Lemmer, M.; Inkpen, M. S.; Kornysheva, K.; Long, N. J.; Albrecht, T., Unsupervised vector-based classification of single-molecule charge transport data. *Nat Commun* **2016,** *7*, 12922.
22. Makk, P.; Tomaszewski, D.; Martinek, J.; Balogh, Z.; Csonka, S.; Wawrzyniak, M.; Frei, M.; Venkataraman, L.; Halbritter, A., Correlation analysis of atomic and single-molecule junction conductance. *ACS Nano* **2012,** *6* (4), 3411-23.
23. Kaliginedi, V.; Rudnev, A. V.; Moreno-Garcia, P.; Baghernejad, M.; Huang, C.; Hong, W.; Wandlowski, T., Promising anchoring groups for single-molecule conductance measurements. *Phys Chem Chem Phys* **2014,** *16* (43), 23529-39.
24. Park, Y. S.; Whalley, A. C.; Kamenetska, M.; Steigerwald, M. L.; Hybertsen, M. S.; Nuckolls, C.; Venkataraman, L., Contact chemistry and single-molecule conductance: a comparison of phosphines, methyl sulfides, and amines. *J Am Chem Soc* **2007,** *129* (51), 15768-9.
25. Hong, W.; Manrique, D. Z.; Moreno-Garcia, 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* (4), 2292-304.
26. Kim, T.; Vazquez, H.; Hybertsen, M. S.; Venkataraman, L., Conductance of molecular junctions formed with silver electrodes. *Nano Lett* **2013,** *13* (7), 3358-64.
27. Yanson, A. I.; Bollinger, G. R.; van den Brom, H. E.; Agraït, N.; van Ruitenbeek, J. M., Formation and manipulation of a metallic wire of single gold atoms. *Nature* **1998,** *395* (6704), 783-785.
28. McNeely, J.; Miller, N.; Pan, X.; Lawson, B.; Kamenetska, M., Angstrom-Scale Ruler Using Single Molecule Conductance Signatures. *The Journal of Physical Chemistry C* **2020,** *124* (24), 13427-13433.
29. French, W. R.; Iacovella, C. R.; Rungger, I.; Souza, A. M.; Sanvito, S.; Cummings, P. T., Atomistic simulations of highly conductive molecular transport junctions under realistic conditions. *Nanoscale* **2013,** *5* (9), 3654-9.
30. Wang, H.; Leng, Y., Gold/Benzenedithiolate/Gold Molecular Junction: A Driven Dynamics Simulation on Structural Evolution and Breaking Force under Pulling. *The Journal of Physical Chemistry C* **2015,** *119* (27), 15216-15223.
31. Friedman, J. H., Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* **2001,** *29* (5).
32. Friedman, J. H.; Meulman, J. J., Multiple additive regression trees with application in epidemiology. *Stat Med* **2003,** *22* (9), 1365-81.
33. He, X.; Pan, J.; Jin, O.; Xu, T.; Liu, B.; Xu, T.; Shi, Y.; Atallah, A.; Herbrich, R.; Bowers, S.; Candela, J. Q., Practical Lessons from Predicting Clicks on Ads at Facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, Association for Computing Machinery: New York, NY, USA, 2014; pp 1-9.

34. Pal, M.; Charan, T. B.; Poriya, A., K-nearest neighbour-based feature selection using hyperspectral data. *Remote Sensing Letters* **2020,** *12* (2), 132-141.
35. Xu, Z.; Huang, G.; Weinberger, K. Q.; Zheng, A. X., Gradient boosted feature selection. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, Association for Computing Machinery: New York, New York, USA, 2014; pp 522-531.
36. Chen, T.; Guestrin, C., XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery: San Francisco, California, USA, 2016; pp 785-794.
37. Altmann, A.; Tolosi, L.; Sander, O.; Lengauer, T., Permutation importance: a corrected feature importance measure. *Bioinformatics* **2010,** *26* (10), 1340-7.
38. Breiman, L., Random Forests. *Machine Learning* **2001,** *45* (1), 5-32.
39. Albanese, D.; Filosi, M.; Visintainer, R.; Riccadonna, S.; Jurman, G.; Furlanello, C., Minerva and minepy: a C engine for the MINE suite and its R, Python and MATLAB wrappers. *Bioinformatics* **2013,** *29* (3), 407-8.
40. Kinney, J. B.; Atwal, G. S., Equitability, mutual information, and the maximal information coefficient. *Proc Natl Acad Sci U S A* **2014,** *111* (9), 3354-9.
41. Reshef, D. N.; Reshef, Y. A.; Finucane, H. K.; Grossman, S. R.; McVean, G.; Turnbaugh, P. J.; Lander, E. S.; Mitzenmacher, M.; Sabeti, P. C., Detecting novel associations in large data sets. *Science* **2011,** *334* (6062), 1518-24.
42. Fu, T.; Zang, Y.; Zou, Q.; Nuckolls, C.; Venkataraman, L., Using Deep Learning to Identify Molecular Junction Characteristics. *Nano Lett* **2020,** *20* (5), 3320-3325.
43. 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* **2020,** *22* (3), 1674-1681.
44. Lauritzen, K. P.; Magyarkuti, A.; Balogh, Z.; Halbritter, A.; Solomon, G. C., Classification of conductance traces with recurrent neural networks. *J Chem Phys* **2018,** *148* (8), 084111.
45. Fatemi, V.; Kamenetska, M.; Neaton, J. B.; Venkataraman, L., Environmental control of single-molecule junction transport. *Nano Lett* **2011,** *11* (5), 1988-92.