Nano-Gap-Based Single-Molecule Measurement for Molecular Discrimination

Jiho Ryu

A thesis submitted to

THE GRADUATE SCHOOL OF SCIENCE

OSAKA UNIVERSITY

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN SCIENCE

February 2024

Contents

Chapter 4. Molecular Modification of Nano-gap Electrodes

Chapter 5. Molecular Discrimination through Chemical Environment Changes

Chapter 6. Direct Single-Molecule Discrimination without Pre-Training

Chapter 1. General Introduction

1.1 Introduction to Single-Molecule Measurement

Single-molecule measurement offers a unique approach to delve into the distinctive characteristics of individual molecules, steering away from the conventional practice of averaging data from ensembles. This technique allows for the examination of the unique properties of individual molecules, departing from traditional ensemble averaging methods. The junction of a molecule with a metal electrode measures the molecule electrically and provides insight into charge transport at the single-molecule level.[1–9]

An early stage of research using single-molecule measurement is the field of molecular electronics, which applies individual molecules to study electronic devices such as transistors[4,10], switches[11,12], and diodes[13–16]. Moreover, the application of this experimental technique provides an opportunity to explore the chemical and physical phenomena at the single-molecule level. The field of single-molecule measurement is expected to lead to significant discoveries across various fields and play a pivotal role in the advancement of science and technology.

The proposal of electrically measuring individual molecules originated in 1974 with A. Aviram and M. A. Ratner, who attempted to construct a rectifier utilizing a single molecule.[17] Since the introduction of this concept, efforts have been made to develop new techniques to measure individual molecules, and today there are many different ways to measure a single molecule. Among these various measurement methods, the break junction approach is most commonly used for connecting individual molecules to metal electrodes.

Break junction refers to a device comprising two metal wires separated by a nanometer-scale gap. There are several platforms for break junction, with Mechanically Controllable Break Junction (MCBJ) and Scanning Tunneling Microscope-Break Junction (STM-BJ) approaches being the most widely utilized. The following sections describe these two break junction

platforms.

1.1.1 Mechanically Controllable Break Junction (MCBJ)

The MCBJ method, initially devised by J. Moreland, J. W. Ekin in 1985[18], and improved by C. Muller, *et al*. in 1992[19,20], offers mechanically-stable electrodes with an adjustable gap. This enables the characterization of single-molecules using uncontaminated atomic nanoelectrodes. MCBJ is one of the most prominent platforms for molecular measurements and can be used to investigate the conducting properties, interactions, and charge transport of molecules. MCBJ is a technique used to precisely manipulate nano-gap by mechanically breaking metal wire bridges, typically composed of materials such as gold or platinum. Figure 1a shows an MCBJ system in its initial state, with a piezoelectrically driven push rod underneath the substrate. As in Figure 1b, when the push rod moves upward, the substrate, anchored in place by the counter supports, undergoes a three-point bend. As the substrate bends increasingly, the metal wire bridge gradually stretches until it eventually breaks, forming a nano-gap. The newly formed nano-gap electrodes are clean, and the nano-gap distance between the electrodes can be controlled by further bending or relaxing the substrate. The distance between the nano-gap electrodes is precisely controlled by the piezo element. The tips of both electrodes can be connected by molecules in the solution to form a molecular junction. This molecular junction is typically formed by the interaction of the molecule's linker groups with the metal electrode. In molecular conductance measurements, the inherent conductance of a molecule is traced by repeatedly forming and breaking molecular junctions. There is also a method for measuring the conductance when molecules freely diffusing in the solution are captured within the nano-gap.

1.1.2 Scanning Tunneling Microscope-Break Junction (STM-BJ)

STM-BJ is a technique that utilizes the nano-space-forming capabilities of a break junction along with a scanning tunneling microscope (STM) to measure the conductance properties of single molecules. The technique, shown in Figure 1.1c, was first reported by Xu and Tao in 2003 and has been in continuous use since then, playing an important role in the field of single-

molecule measurement.[21]

In STM-BJ experiments, the STM tip is precisely controlled on the metal substrate at the angstrom level to form a metal nanocontact. Then, the tip is inserted until only a short chain of gold atoms remains. During this process, the conductance of the gold contact formed between the tip and the metal substrate exhibits a stepwise decrease at integer multiples of G_0 ($=2e^2/h$) as the tip moves away from the substrate. If the metal contact is broken in the presence of molecules in solution, the molecules can fill the gap between the tip and the substrate. The formation of a stable molecular junction between the tip and the substrate results in the observation of characteristic conductance. Conversely, in the absence of molecules in the solution, no characteristic conductance is observed. This behavior underscores the sensitivity and specificity of the STM-BJ technique in detecting molecular interactions within nanoscale gaps.

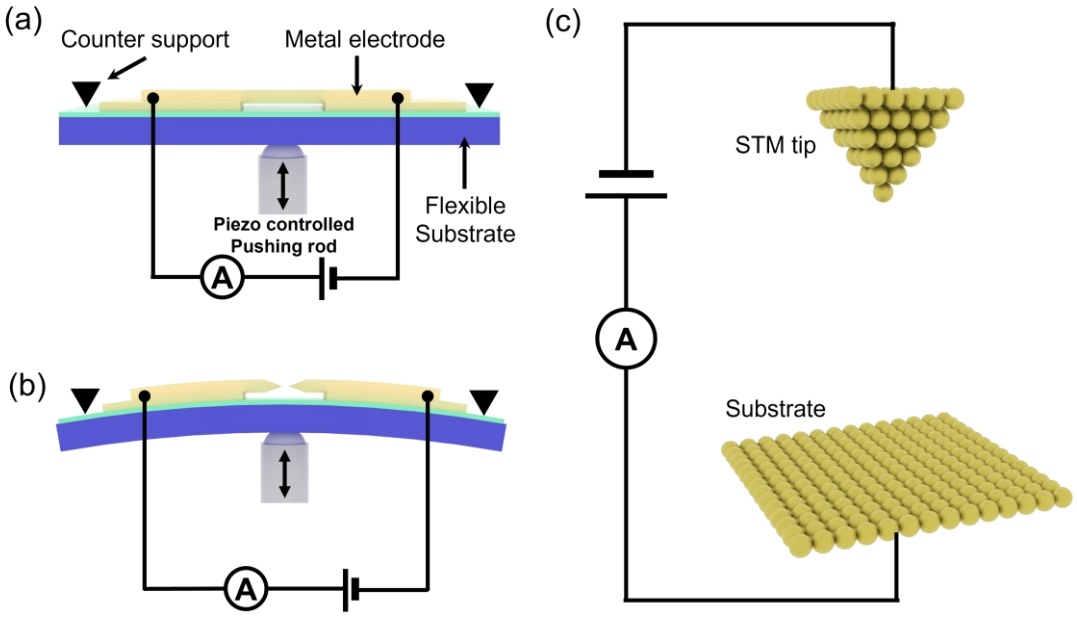

Figure 1.1 Schematics of MCBJ and STM-BJ systems. (a), (b) illustrate MCBJ, and (c) illustrate STM-BJ. (a) The initial state of the MCBJ system, where the pushing rod is at the origin and there is no bending of the substrate. (b) The substrate is bent due to three-point bending caused by the elevation of the pushrod. The nano-wire is broken and a nano-gap is formed.

1.2 Investigation of Chemical/Physical Phenomena

In the preceding section, the potential of single-molecule measurements and the experimental techniques were explored. Single-molecule measurement is a technique utilizing metal nano-gap electrodes to measure individual molecules and analyze their electrical conductance.[1–9] This section describes the chemical and physical process investigated by single-molecule measurements. The observation and understanding of molecular behavior at the single-molecule level contribute significantly to the exploration of chemical, physical events at the nanoscale, with these measurements playing a key role.[22–27] This section covers the importance of single-molecule measurements with an introduction to some studies.

A. C. Aragones, *et al*. reported that electrostatic effects can influence chemical reactions. This study provided experimental evidence demonstrating the acceleration of chemical reaction in the presence of electric fields.[28] During experiments with STM-BJ, the accelerated formation of carbon-carbon bonds under the influence of an electric field was confirmed in the Diels-Alder reaction. These findings hold the potential to open avenues for innovative approaches to chemical catalysis. C. Yang, *et al*. reported identifying the actual mechanism among the two anticipated mechanisms for the Suzuki-Miyaura cross-coupling reaction using a high-resolution single-molecule detection platform.[29] The study employed the platform to elucidate the entire reaction pathway, providing detailed insights. By detecting the electrical signals associated with oxidative addition/ligand exchange and metal transfer, the researchers clarified the metal exchange mechanism. This was accomplished using a single-molecule Pd catalyst integrated on a nano-gap graphene point electrode. Additionally, the study determined the kinetic constants for each step and the overall catalytic time scale.

Additionally, single-molecule measurement experiments employing a variety of molecules have provided insights into diverse physical events within molecules. M. Taniguchi, *et al*. used a naphthalenedithiol to investigate changes in the conductance of molecular junctions based on molecular symmetry.[30] The 1,4-symmetric molecular junction, in comparison to its 2,7-symmetric counterpart, exhibited conductance 110 times higher, offering a detailed theoretical and experimental understanding of the impact of molecular symmetry on conductance. These findings showed the potential to design high-conductance molecular systems through the control of molecular symmetry.

The deep understanding of chemical and physical events through single-molecule measurements has shown the potential of measurements at the molecular scale in exploring new science and technology. The field of single-molecule measurement is expected to provide new insights in various fields such as medicine and biology.

1.3 Application to Bio-Molecular Analysis

In addition to the research areas discussed in Section 1.2, several other significant fields have been explored using single-molecule measurements. Outstandingly among various fields, the analysis of biomolecules receives notable attention. Single-molecule measurements enable direct detection of molecules, obviating the need for sample processing and allowing for simple and rapid analysis. Moreover, theoretically capable of detecting a molecule, it possesses the potential for analysis at extremely low concentrations. Using the advantages of single-molecule measurements, diverse analytical methods for biomolecules, including RNA/DNA[31–36], amino acids[37], peptides[37,38], proteins[39–41], neurotransmitters[42], sugars[43].

Polymerase chain reaction (PCR) is a molecular biology technique for replicating and amplifying desired portions of DNA. It is currently the most widely used technique for DNA sequencing, but it requires DNA amplification, which is inherently time-consuming and expensive. The ultimate method of DNA/RNA sequencing is to read each nucleotide group from a single piece of DNA/RNA and analyze the sequence simultaneously. The conceptualization of these technologies began in 2008 with a theoretical proposal by Di Ventra group for DNA/RNA sequencing via single-molecule measurements.[44–46] Figure 1.2 represents their proposed concept of direct DNA/RNA sequencing. The principle of this sequencing technique is that when DNA is passed between appropriately spaced electrodes, each nucleotide generates a unique current signal due to the unique chemical structure of the base. Therefore, by passing DNA between electrodes and analyzing the current generated over time, its sequence can be determined. Following the theoretical confirmation of the feasibility of single-molecule DNA/RNA sequencing, there has been a significant increase in research interest focused on analyzing DNA/RNA at the single-molecule level.

H. Tanaka and T. Kawai accomplished a breakthrough in DNA sequencing on a copper

surface utilizing a Scanning Tunneling Microscope (STM).[47] By employing an oblique pulse-injection method to deposit molecules onto the surface, they overcame challenges associated with preparing samples of long-chain DNA molecules. The study revealed a distinct electronic state of guanine bases, enabling the sequencing of individual guanines. Through high-resolution scanning tunneling microscope imaging and spectroscopy, the researchers identified the 'electronic fingerprint' of guanine bases in the DNA molecule, demonstrating the possibility of sequencing individual guanine bases in real long-chain DNA molecules. This work represents a significant advancement beyond the mere visualization of DNA and showcases the potential to distinguish and sequence individual DNA nucleobases.

T. Ohshiro, *et al*. reported a groundbreaking single-molecule electro-randomized sequencing of DNA and RNA, employing tunneling current and randomized sequencing.[34] As shown in Figure 1.3, individual DNA bases in aqueous solution were measured and analyzed by MCBJ. This approach facilitated the identification of different combinations of DNA oligomers, yielding conductance-time profiles. These profiles were then employed to distinguish the base molecules within each oligomeric configuration, enabling the resequencing of various DNA oligomers and miRNAs, as reported. These studies represent significant strides in DNA/RNA sequencing techniques at the single-molecule level.

Single-molecule measurements can directly measure a target molecule without additional sample processing steps. This advantage facilitates faster and more sensitive analysis and detection of various molecules. Given its high potential for biomolecular analysis, single-molecule level biomolecular analysis has been applied across a diverse range of biomolecule families, including proteins/peptides, amino acids, neurotransmitters, and more. In the pursuit of developing peptide sequencing techniques, T. Ohshiro, *et al*. conducted measurements and discrimination of amino acids.[37] As Figure 1.4 shows, this study attempted to discriminate between amino acid molecules using the current properties of duration and conductance. This data indicates that while some amino acids are discriminable, accurate discrimination is not easy due to similarities in current properties.

The overlap in current properties originates from the similarity in electronic structure, and these molecules may possess similar molecular structures. Therefore, in the field of single-molecule measurement, discriminating between molecules with similar current properties

poses a significant challenge. Since histogram-based analysis methods only provide statistical conductance information, molecular discrimination becomes difficult when histograms overlap. In particular, currents from single-molecule measurements have large variations, so histogram-based analysis may not be an appropriate method for molecular discrimination. To overcome these limitations and improve the accuracy of molecular discrimination, a machine learning-based approach has been applied to analyze individual current pulses through single-molecule measurements. The upcoming section will cover the utilization of machine learning. In the following section, the application of machine learning is discussed in detail.

Figure 1.2 Schematic illustration of sequencing for DNA (or RNA) passing through nano-gap electrodes. The colors of each DNA strand represent the four nucleotides, and the color does not signify a specific molecule. As the DNA passes between the nano-gap electrodes, a current is generated by each nucleotide in the DNA. Each nucleotide exhibits distinct electrical characteristics based on its specific chemical structure. Therefore, by analyzing the current over time, DNA/RNA can be sequenced.

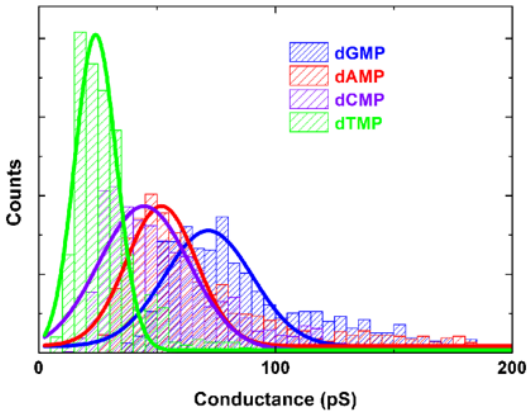

Figure 1.3 Conductance histograms for the four deoxynucleotides of DNA, each containing 1000 maximum current values for the signal of every measured molecule. The blue, red, purple, and green lines represent the single-molecule conductance of dGMP, dAMP, dCMP, and dTMP, respectively. The color code of each line corresponds to the Gaussian fit of the histogram. Figure reprinted from "Single-Molecule Electrical Random Resequencing of DNA and RNA" by T. Ohshiro, *et al*., 2012, Article, *2*(1), 501, Copyright 2012 by Springer Nature.

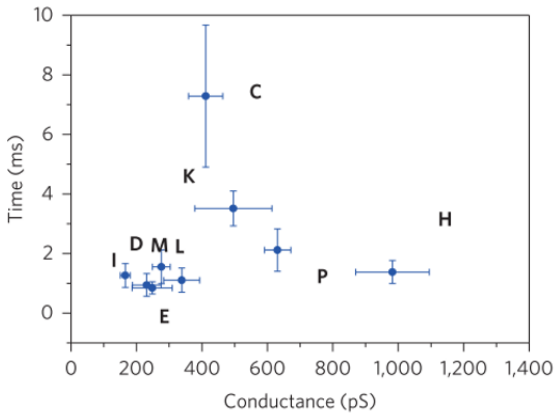

Figure 1.4 The two-dimensional scatter plot of conductance and duration from single-molecule measurements of amino acids diffusing in solution. Each data point represents Aspartic acid (D), leucine (L), histidine (H), proline (P), glutamic acid (E), isoleucine (I), lysine (K), cysteine (C), methionine (M). Figure reprinted from "Detection of post-translational modifications in single peptides using electron tunnelling currents" by T. Ohshiro, *et al*., 2014, Letters, *9*(10), 835-840, Copyright 2014 by Springer Nature, and used with permission.

1.4 Application of Machine Learning

In the field of single-molecule measurement, histograms have traditionally been used as the primary tool for analyzing measured data and determining the conductance of molecules. However, as discussed in the preceding section, single-molecule currents exhibit significant variability, posing challenges when attempting to discriminate between molecules. The experiments in this thesis, molecules of different orientations are measured as they diffuse freely in solution, without chemically bonding to nano-gap electrodes, thereby introducing even greater variability into the data. Figure 1.4 in the previous section shows single-molecule measurement data depicting various orientations of an amino acid molecule diffusing freely in solution, with error bars on each data point representing the variability in single-molecule conductance. This means discriminating between different molecules becomes challenging when their current properties overlap. While it may be possible, in some instances, to discriminate based solely on conductance-time data if the molecules exhibit clear differences in current characteristics, molecules of high discrimination importance often share similar current characteristics due to structural similarities.

The application of machine learning has been proposed as an alternative to address the limitations associated with traditional analysis methods. Traditional analysis methods, such as the histogram and maximum current and duration scatter plot in the previous section, do not include detailed information about individual current pulses. Recent advances in machine learning have highlighted its potential applications in various fields such as chemistry and nanoscience.[48–51] Machine learning has also been applied to the field of single-molecule measurements, where it has proven effective in analyzing the current properties of single molecules.[42,45,52–59] For instance, Y. Komoto, *et al*. conducted measurements on three neurotransmitter molecules and applied machine learning to analyze individual current signals for discriminating the target molecules.[42] Figure 1.5a schematically illustrates the measurement of molecules using nano-gap electrodes. Figure 1.5b displays the structures of the three molecules, while Figure 1.5c presents the confusion matrix representing the discrimination accuracy of these molecules. In the confusion matrix, the true sample axis pertains to the target sample actually measured, and the predicted sample axis reflects the degree of prediction for each target molecule by machine learning. This study underscores that the discrimination

accuracy of molecules can be significantly enhanced by employing machine learning for multidimensional analysis of individual single-molecule current data.

The approach of analyzing individual single-molecule current signals through machine learning, capable of handling multidimensional data, represents a novel strategy for discriminating molecules that prove challenging for traditional analysis methods. This method marks a substantial milestone in the advancement of molecular discrimination technology and is expected to broaden the applicability of single-molecule measurements.

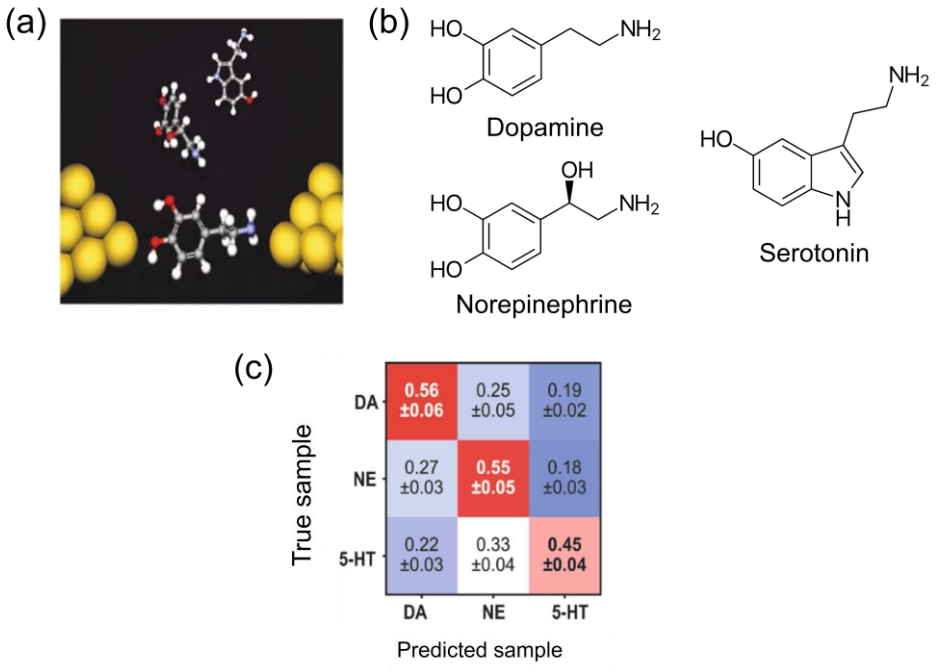

Figure 1.5 Machine learning discrimination results from single-molecule measurement data of three molecules. (a) Schematic of single-molecule measurement. (b) Three target molecules. (c) Confusion matrix showing discrimination results. True sample axis refers to the actual measured sample, Predicted sample axis refers to the molecules predicted by machine learning. DA, NE, and 5-HT stand for dopamine, norepinephrine, and serotonin, respectively. Figure reprinted from "time-resolved neurotransmitter detection in mouse brain tissue using an artificial intelligence-nanogap" by Y. Komoto, *et al*., 2020, Article, *10*(1), 11244, Copyright 2020 by Springer Nature.

1.5 To broaden the Application Scope of Single-Molecule Measurement

In the previous sections, the various scientific discoveries and potential applications of single-molecule measurements across a wide range of fields were discussed. A notable aspect is the enhancement of the potential of single-molecule measurements due to the application of machine learning. In the case of molecular discrimination, machine learning classifiers are trained on measurement data of pure target molecules and then applied to the analysis of actual samples. Therefore, the improvement of single-molecule discrimination accuracy is strongly required for the practical application of molecular discrimination. Many molecules that are important to discriminate have similar molecular and electronic structures, resulting in similar current properties, making discrimination difficult. To address these issues, my main goal in this thesis is to develop a novel approach to improve the discrimination accuracy of molecules that are difficult to discriminate by conventional methods due to their similarity in molecular structure. I will also present and experimentally validate an innovative approach to apply single-molecule measurement-based molecular discrimination methods to a wider range of molecules.

In this thesis, I propose two approaches to improve molecular discrimination accuracy. The first approach involves modifying nano-gap electrodes with molecules. By utilizing suitable modifying molecules, it is anticipated to interact with the target molecules, inducing a change in its current properties. These interaction differences can be a factor in effectively improving molecular discrimination accuracy. The second approach induces behavioral differences in target molecules through changes in chemical environment. Molecules with similar current properties under normal conditions may exhibit different behaviors under chemical environment changes, which can improve the accuracy of molecular discrimination. These experiments are also expected to provide insight into the unique properties of single-molecules by observing differences in molecular behavior in response to changes in the chemical warfare environment. As a result, these two methodologies will lay the foundation for new approaches that can significantly improve the discrimination accuracy of various molecule.

In addition, I propose a new analytical method that can further broaden the application of single-molecule measurements. Considering the practical application of single-molecule measurement for the detection of specific molecules, it is probably not easy to prepare and train

pure solutions for all molecules in a sample. Therefore, the development of a method that can discriminate molecules directly in a mixed sample may represent a major advance in the field of single-molecule measurement. These advances in molecular discrimination approaches are expected to greatly accelerate the adoption of single-molecule measurements in a variety of applications.

Chapter 2. Fundamentals

The advancement of single-molecule measurement techniques has led to a profound exploration of chemical and physical processes at the molecular level, contributing significantly to gaining innovative scientific insights across various fields. While electron transport through molecules between nano-gap electrodes is easily conceivable, a precise understanding of this phenomenon remains unclear. Nevertheless, previous research has proposed models and established some concepts and techniques to elucidate charge transport through molecules. Additionally, this study applies machine learning to analyze single-molecule measurement data, departing from traditional histogram-based methods. This chapter delves into the models elucidating the movement of electrons through single molecules and provides details about the machine learning algorithms employed in this research.

2.1 Models for Describing Transmission through Molecule

The tunneling phenomenon is a quantum mechanical effect wherein quantum particles, such as electrons, can traverse barriers even at low energies, defying classical mechanics' prediction of their inability to pass through. In classical mechanics, it is considered impossible for a particle to surpass a barrier with higher energy than the particle itself. However, in the microscopic realm governed by quantum mechanics, particles like electrons exhibit tunneling through barriers, a phenomenon attributed to the wave nature of matter. The probability of transmission through tunneling can significantly vary based on conditions such as the height and width of the barrier and the mass of the particle. This study focuses on single-molecules, presenting a model to elucidate electron transport through individual molecules.

In 1957, R. Landauer proposed a formula for transmission in cases where the size of system becomes sufficiently small, necessitating the consideration of the wave nature of electrons.[60] According to Landauer, determining the conductance in a system involves identifying the intrinsic modes through which current flows, calculating transmission values, and aggregating their contributions. This process is mathematically summarized by the Landauer formula as

follows.

$$G = \frac{2e^2}{h}\sum_{n=1}^{N} T_n \qquad Equation\ 2.1$$

Where electrical conductance (G) signifies the overall capacity for electrical conduction, while e^2/h represents the quantum unit of conductance determined by the electron charge (e) and the Planck constant (h). T_n refers to the transmission of an individual transition mode. In this context, $G_0=2e^2/h$ represents the fundamental unit for quantifying conductance, emphasizing the quantum nature of electrons and playing a crucial role in explaining the behavior of electrons in extremely small systems. The Landauer formula underscores the quantum characteristics of electrons in minimized systems through the introduction of this foundational conductance unit, G_0.

As elucidated by the Landauer formula, electron transmission in highly restricted spaces delves into the realm of quantum mechanics. Specifically, the phenomenon of electrons tunneling through potential energy barriers in quantum mechanics exerts a substantial influence on conductance, akin to the conceptualization of electrons effectively penetrating energy barriers. A pivotal feature of the Landauer formula is its depiction of conductance (G) through transmission rates (T_n). Transmission rate is construed as the probability of electron transmission in a specific transition mode. Consequently, electron transmission in small spaces involves tunneling through various transition modes, with each mode's transmission rate contributing to the overall conductance. These quantum mechanical attributes play a critical role in determining the comprehensive characteristics of electron transmission. Notably, the structure of the system, potential energy barriers, and molecular-level intricacies affect the tunneling probability of electrons. When a molecule is introduced between metal nano-gap electrodes, the transmission of electrons can be assessed using the Breit-Wigner formula.[61] This formula is intended to model the efficiency of the electron transmission process at specific energy levels.

$$T(E) = \frac{4\Gamma_L\Gamma_R}{[E-E_0]^2+[\Gamma_L+\Gamma_R]^2} \qquad Equation\ 2.2$$

Where Γ denotes the coupling strength between the molecule and the metal, while $_L$ and $_R$

represent the coupling interactions with the left and right electrodes, respectively. E represents the energy of the tunneling electron, and E_0 indicates the energy level of the molecular orbital. According to the Breit-Wigner formula, the transport of electrons through the molecule is influenced by the alignment of molecular conduction orbitals interacting with the electrodes through coupling constants Γ_L and Γ_R. In simple terms, the Breit-Wigner formula suggests that charge transport through a molecule is more efficient when the energy levels of the molecular orbitals closely match the Fermi energy and when the bond between the molecule and the electrode is stronger. Figure 2.1 shows a model of the energy level of an electrode at zero bias. When a bias voltage is applied, the anode moves energy downward and the cathode moves energy upward. Then, the potential difference between the two electrodes allows current to flow. The Breit-Wigner curve between the two electrodes represents the molecular orbital level and indicates the probability of electron transport through the molecule. As the conduction band levels of the molecule approach the Fermi level of the electrode, transmission becomes dominant, and the transmission probability correlates with the area under the curve. This region symmetrically envelops the Fermi energy level, with its breadth modulated by the applied bias voltage. The Breit-Wigner formula underscores that the transmission efficiency in a single-molecule system is not a fixed value but can vary based on factors such as the molecular structure, the strength of coupling between the molecule and electrodes, and the applied voltage.

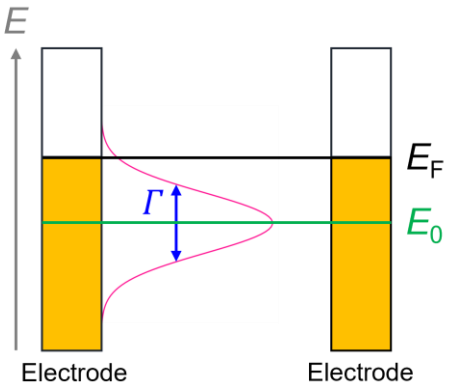

Figure 2.1 Energy level (E) model for describing the Breit-Wigner formula in single-molecule measurements. The symbols are based on the Breit-Wigner formula (Equation 2.2). E_0 represents the energy level of conduction orbital, E_F represents Fermi level of metal electrode, Γ represents the coupling strength between molecular and electrodes.

2.2 Machine Learning Algorithms for Classification

The integration of nanotechnology and data science has facilitated an in-depth comprehension of molecular behavior through precise electrical measurements at the single-molecule level, coupled with meticulous data analysis. In this research, individual molecules in solution underwent electrical measurements as they traversed between gold nano-gap electrodes. The resulting data underwent analysis employing machine learning techniques to discern between distinctive current signals. The machine learning classifier was trained using features, numerical attributes extracted from the current characteristics of a molecule based on user-defined criteria, and subsequently utilized the training data for classification on new data. This chapter delves into the discussion of the Random Forest and XGBoost algorithms applied in this study.

2.2.1 Random Forest

Random Forest was initially proposed by L. Breiman and A. Cutler in 2001.[62] Known as an ensemble learning method, Random Forest generates multiple decision trees, combining them to reduce overfitting and create a stable predictive model. Each decision tree is created using different subsets of data and randomly selected features. While each tree operates independently, Random Forest performs predictions collectively, leveraging their combination for enhanced predictive capabilities. This contributes to improving model stability and effectively overcoming overfitting issues.

In Random Forest, the GINI index is employed to partition the data. When a single node is split, it may not distinctly separate all data based on the node's conditions. The degree of non-separation at this node is referred to as impurity. The increase in impurity at each node in the decision tree results in less purity, potentially leading to overfitting to the training data and a decrease in predictive performance on new data. In Random Forest, the GINI index represented by Equation 2.3 is used for data partitioning. This GINI index serves as a metric to measure impurity in each node's data split.

$$GINI\ index = 1 - \sum_{i=1}^{n} (p_i)^2 = 1 - \left[{P_+}^2 + {P_-}^2\right] \qquad Equation\ 2.3$$

where P_+ represents the probability of a positive class, and P_- signifies the probability of a negative class. Random Forest splits the nodes of each tree in a way that minimizes the GINI index, aiming to form a robust predictive model by combining the predictions of each tree. Each tree is constructed using a combination of bootstrapped sampling and random feature selection, fostering the creation of diverse trees. Bootstrapped sampling, which allows for the selection of data with replacement, ensures that each tree employs a distinct subset of data. Random feature selection involves considering only a subset of features at each node, encouraging each tree to focus on different features and thereby increasing diversity. These elements secure diversity in each tree, enhancing the model's predictive performance and addressing overfitting effectively.

Random Forest employs a combination of bootstrapped sampling and random feature selection to form diverse trees. The predictions of all generated decision trees are then aggregated, utilizing a majority voting approach. The Random Forest algorithm illustrated in Figure 2.2 showcases robust predictive performance and stability, making it widely applicable in various fields, particularly excelling in handling high-dimensional data and diverse features. The effectiveness of Random Forest in high-dimensional data can be attributed to its flexible variable selection in such datasets, where numerous features may be present.

Firstly, Random Forest exhibits flexibility in variable selection in high-dimensional data. Randomly selecting features during the construction of each decision tree allows the algorithm to focus on different features for each tree. This concentration on diverse features aids in emphasizing important characteristics while disregarding unnecessary ones, thereby enhancing the model's performance. Secondly, Random Forest enhances diversity using bootstrapped sampling and random feature selection. Bootstrapped sampling ensures that each tree operates on a different subset of data, contributing to increased model diversity. Random feature selection, considering only a subset of features at each node, aids in forming decision trees that account for various features. These features enable Random Forest to effectively capture diverse and intricate interactions present in high-dimensional data, resulting in stable and

robust performance.

Random Forest proves to be an excellent choice for handling high-dimensional data and diverse features. In this study, given the transformation of measured individual single-molecule current pulses into a 13-dimensional feature space, Random Forest was adopted as an algorithm capable of managing high-dimensional data.

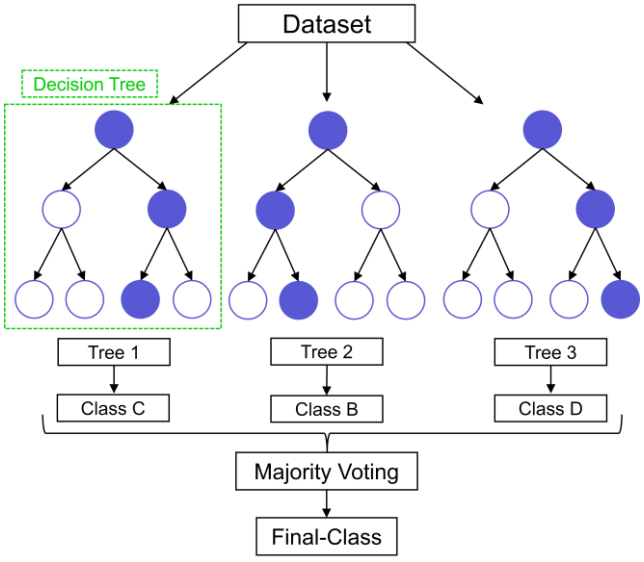

Figure 2.2 The illustration of Random Forest Algorithm.

2.2.2 XGBoost

XGBoost, short for Extreme Gradient Boosting, is a powerful machine learning algorithm widely employed in recent data science applications. XGBoost was introduced by T. Chen in 2016 and is part of the family of gradient-boosted decision trees (ensembles), offering enhanced performance and speed via a sequential decision tree-based machine learning approach.[63] It offers parallel tree boosting capabilities and can be applied to various machine learning problems such as regression, classification, and ranking. Particularly well-suited for large datasets and complex prediction tasks, XGBoost is frequently utilized due to its accuracy and execution speed.

Key features of XGBoost include the Gradient Boosting technique and a regularization approach. Boosting is an ensemble learning technique that combines weak learners to create a strong learner. Weak learners are either randomly selected or weighted based on the errors of previous learners. Boosting involves sequentially training multiple weak learners, allowing each model to correct the errors of the previous model. Among boosting techniques, Gradient Boosting utilizes Gradient information, meaning each model is trained using Gradient information related to the errors of the previous model. This helps the model reduce the prediction errors of the previous model, enhancing predictive capabilities.

One crucial fundamental concept used in XGBoost is CART (Classification and Regression Trees). CART is a tree-based learning algorithm for solving classification and regression problems. It constructs a tree using a recursive split approach based on features at each node, and the resulting tree is used to predict classes at each leaf node. XGBoost leverages CART as the foundation to form a robust predictive model. XGBoost trains the model by optimizing the objective function, represented by Equation 2.x. This objective function comprises a loss function and a regularization term, enhancing the model's predictive ability and preventing overfitting.

$$obj(\theta) = \sum_{i}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{k} \Omega(f_k) \qquad Equation\ 2.4$$

Where $\hat{y}$ is the model's predicted value, y is the model's true value, Ω is the regularization term, and f_k is each individual tree. In XGBoost, the objective function (θ) consists of a function that calculates the loss between the actual values and the model's predictions for each individual data point i, along with a regularization term imposing complexity on each individual tree. Consequently, through the objective function, XGBoost prevents overfitting and enables more generalized predictions. Figure 2.3 illustrates a schematic representation of XGBoost.

Particularly, XGBoost exhibits excellent performance in high-dimensional data analysis, attributed to several aspects. Firstly, it demonstrates flexibility in variable selection. In high-dimensional data, numerous features may be present. XGBoost calculates the importance of each feature when constructing each tree, focusing on significant features to form the tree. This

enables the model to ignore unnecessary features, reducing complexity. Secondly, the objective function of XGBoost includes a regularization term controlling the complexity of each tree, preventing an excessive reliance on specific features. This regularization is particularly beneficial for enhancing the generalization performance of the model in high-dimensional data. Thirdly, XGBoost possesses the robustness of Gradient Boosting. Each tree is trained to correct the prediction errors of the previous tree, allowing XGBoost to capture complex patterns or interactions prevalent in high-dimensional data. Lastly, XGBoost allows parallel processing and optimization, efficiently operating on large datasets, especially suited for handling substantial processing in high-dimensional data. These features collectively empower XGBoost to demonstrate outstanding predictive performance in high-dimensional data.

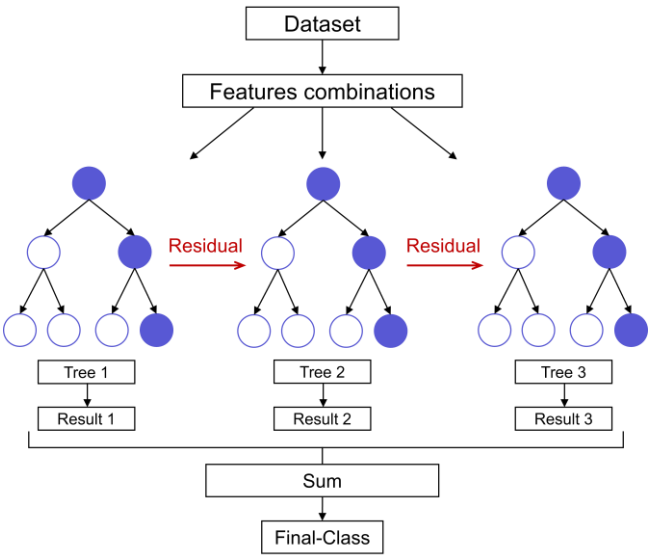

Figure 2.3 The illustration of XGBoost Algorithm.

2.2.3 F-measure in Machine Learning Classification

The F-measure, or F-score, is a crucial metric employed in evaluating the performance of machine learning models. It is calculated as the harmonic mean of precision and recall, providing an overall assessment of the model's performance, particularly valuable in scenarios with imbalanced class distributions. Precision represents the proportion of instances predicted

as positive by the model that are indeed positive and is calculated by the formula below.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \qquad Equation\ 2.5$$

Precision measures the accuracy of predictions for all classes, but it may not be an appropriate evaluation metric in cases of severe class imbalance. Recall, as expressed by the formula below, represents the proportion of actual positives that the model accurately predicts as positive..

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \qquad Equation\ 2.6$$

A high recall implies that the model effectively detects true positives without missing them. However, as recall increases, precision may decrease, illustrating a tradeoff relationship with precision, which represents the proportion of accurately predicted positives. This suggests the difficulty of simultaneously improving precision and recall. Elevating precision enhances the ratio of correctly predicted positive samples but increases the number of missed positive samples, leading to reduced recall. Conversely, increasing recall identifies more positive instances but can result in decreased precision.

Due to the trade-off relationship between precision and recall, evaluating performance with a single metric can be challenging. To address this challenge, the F-measure is introduced. The F-measure, calculated as the harmonic mean of precision and recall, serves as an evaluation metric providing a balanced assessment between the two metrics and is computed as follows.

$$F - measure = 2\ \frac{Precision * Recall}{Precision + Recall} \qquad Equation\ 2.7$$

The F-measure provides a comprehensive evaluation of the predictive capability of the positive class, assessing how well the model identifies and accurately predicts positive samples. In essence, the F-measure is a valuable metric that goes beyond accuracy, enabling a more holistic understanding of the model's predictive capabilities.

Chapter 3. Methods

In this study, gold nano-gap electrodes formed through MCBJ were utilized for the electrical measurement of single molecules. Subsequently, the obtained data were employed to train a machine learning classifier, enabling the prediction of the mixing ratio of compounds. This chapter covers the fabrication process of the nano-gap device, formation of the nano-gap, measurement process of single molecules, and the data analysis procedure utilizing machine learning.

3.1 Fabrication of Nano-gap Device

In this study, MCBJs were used as a platform to form nano-gap electrodes and measure single molecules. To form nano-gap electrodes using the MCBJ system, a flexible substrate-based device with metal nano-wires is required. The nano-gap device used in this study was fabricated via micro/nano fabrication.

This section describes the fabrication process of the MCBJ nano-gap device used in this study. Figure 3.1 shows a schematic of the fabrication process of the nano-gap device. First, polyimide is spin-coated as an insulating layer on a flexible Si substrate with a thickness of about 100 μm and baked. Then, the electrodes of the MCBJ measurement device, the connecting pad of the nano-gap device, and other relatively large-sized electrodes are formed using photolithography on the polyimide. At this time, the center part of the nano-gap device was not lithographed. The next step was to form tens of nanometer-wide gold nanowire in the center of the substrate using electron beam lithography. Then, SiO2 was deposited using chemical vapor deposition to passivate the exposed areas of the gold electrode for low-noise tunnelling current measurements for single-molecule detection in an electrolyte solution using insulator-protected nanoelectrodes.[64,65] Figure 3.2 shows the actual fabricated nano-gap device and the SEM image of the sensing part in the center.

In this study, single-molecules are measured in solution, so the nano-gap device needs a

separate device to contain the solution. As shown in figure 3.3a, a polydimethylsiloxane (PDMS) well was fabricated and attached to the top of the chip to contain the sample solution. The junction of the MCBJ chip and the PDMS well was then treated in O2 plasma for 10 seconds and then placed in a vacuum oven at 90°C for 60 minutes. Evaporation of the solvent will change the concentration of the sample, so to prevent this, immediately after the sample was injected into the PSMS well, Kapton tape was applied to seal the sample from the outside environment, as shown in Figure 3.3b.

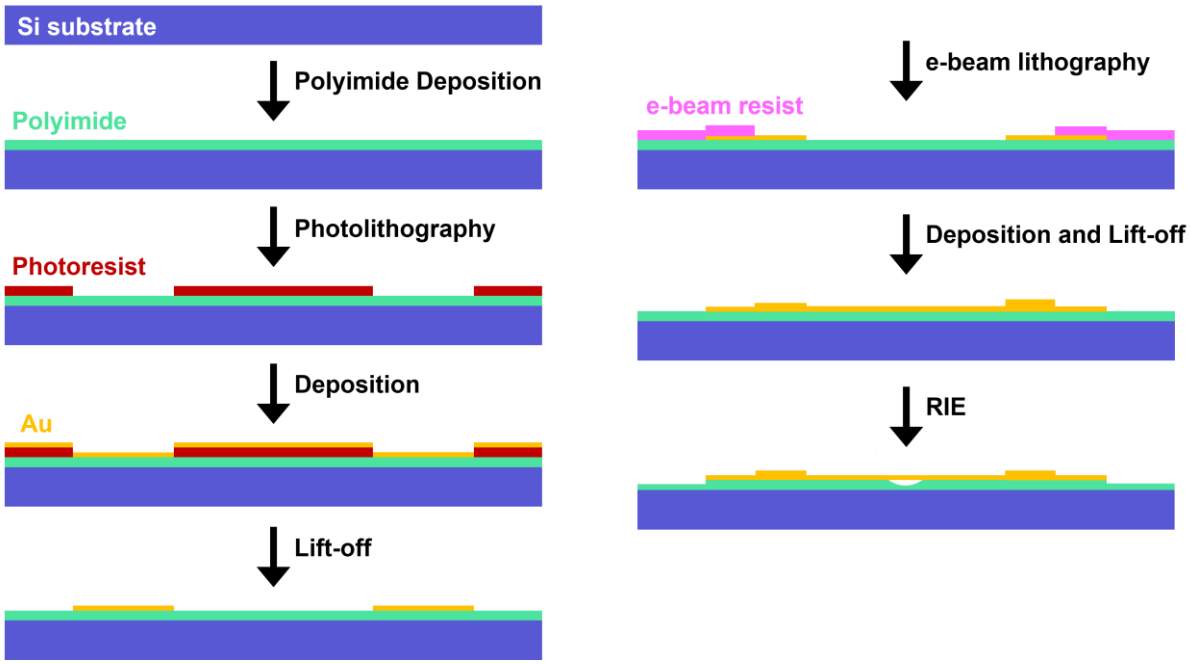

Figure 3.1 Schematic of the fabrication process for MCBJ nano-gap devices. The Si substrate is flexible and about 100 um thick. The polyimide was cured at 250°C. After each lithography, 5 nm Cr is deposited as a step prior to gold deposition. This illustration depicts the side view of the long edge, revealing the cross-section cut from the center of the short side.

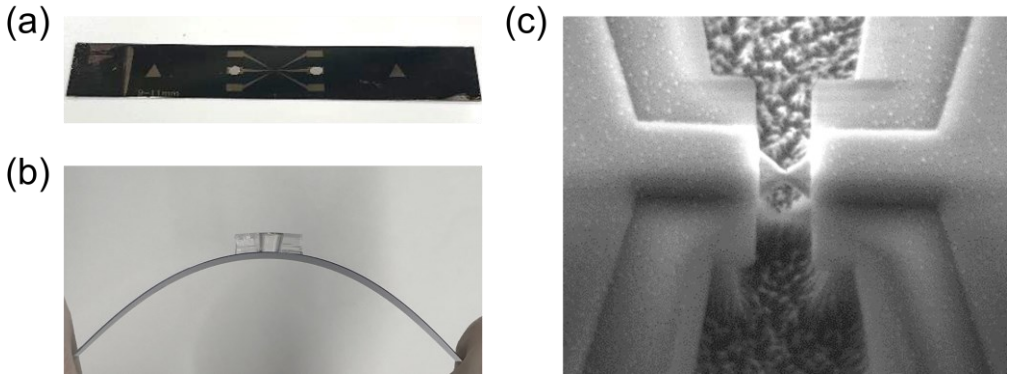

Figure 3.2 Images and SEM image of nano-gap device for MCBJ. (a) Top image of the nano-gap device. (b) side view image of a bent nano-gap device with a PDMS well attached. (c) SEM image of the sensing part of the nano-gap device. The ribbon-shaped part in the center is broken to create a nano-gap.

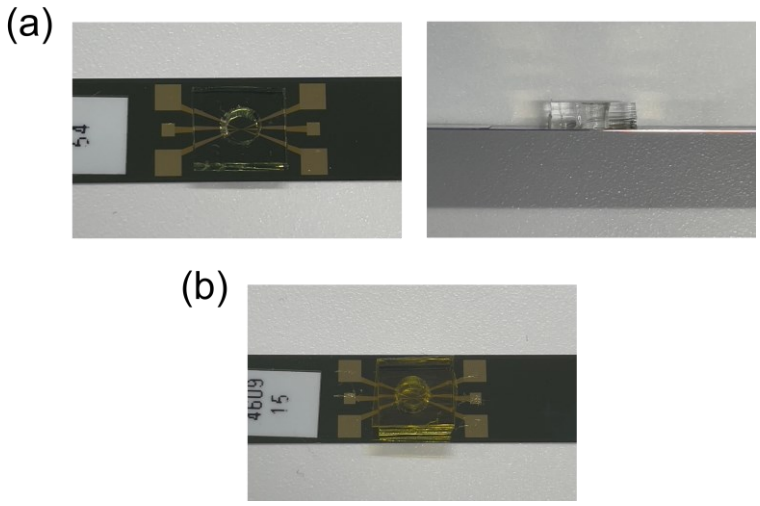

Figure 3.3 Images of the MCBJ device chip with PDMS well attached. (a) Top and side images. Circled holes in the square PDMS sheet. Thickness is approximately 5 mm. (b) Solution injected into PDMS wells and closed with Kapton tape.

3.2 Process of Single-Molecule Measurement

Upon injection of the sample into the nano-gap device and completion of all measurement preparations, the chip is inserted into the MCBJ measurement device for analysis. Key components of the MCBJ measurement system comprise a PC (NI-PXIe), picoammeter (K6487), MCBJ box, control box, and piezo driver. In Figure 3.4, the interconnectedness and nomenclature of these components are depicted schematically. The PC assumes overall control responsibilities, while the control box, tethered to the PC, regulates the entire system, encompassing the stepping motor and piezo driver. The latter is instrumental in manipulating the nano-gap distance by actuating the pushing rod for nano-gap bending control. The picoammeter, integrated with the control box, facilitates requisite current measurements. It monitors current flow throughout the nano-gap system during all experimental phases, from setup to actual data collection. With a resolution of 0.01 pA, the picoammeter can measure currents up to 20 mA.

Principal experimental apparatus is housed within the Faraday box, the locus of all experimental operations. Herein lies a DC power source that furnishes low-noise voltage during data collection, alongside relays that manage system configurations across different stages. Additionally, the Faraday box incorporates a stepping motor and a piezo motor, employed for nano-gap chip manipulation through varying step sizes. Centrally positioned within the Faraday cage is a jig, designed to secure the nano-gap device in place.

Figure 3.5a is an image of the device actually used in the experiment. The nano-gap device is inserted into the jig within the Faraday box shown in figure 3.5b. Figure 3.5c shows the nano-gap device inserted into the jig, and Figure 3.5d shows the jig covered. The lid of the jig has electrodes that can contact the pads on either side of the nano-gap device. The lid also has counter supports on either side of the lid to hold the nano-gap device in place. Figure 3.5e shows an image of the nano-gap device in a bent state with three-point bending by a pushing rod.

After preparing the nano-gap device, the measurement procedure begins by inserting it into the Faraday box. In the first step, the nano-gap formation step, the piezo controlled pushing rod is slowly raised to a set height. As the nano-gap device is subjected to three-point bending by

the pushing rod, the nano-wire continues to become thinner, which reduces the current value. After repeated bending, the nano-wire mechanically breaks, causing the current value to drop sharply, confirming the formation of a nano-gap. With these steps, one nano-gap formation is completed, and the reproducibility of two additional nano-gap formations is checked. If the nano-wire does not break at the initially set height, continue to observe the formation of the nano-gap while gradually increasing the maximum height of the pushing rod.

After confirming the formation of the nano-gap, single-molecule measurement was performed after various checks and training procedures between the pushing rod and the nano-gap electrodes. After the nano-gap was formed, a series of training procedures were performed to properly control the nano-gap distance by the voltage of the piezo drive. The process involves training the cut and reconnect positions by repeatedly raising and lowering the pushing rod. Next, we go through the cut mode to make a clean and pointed nano-gap, and finally, we apply various types of bias voltage to see the current rise. Next, set the nano-gap distance and measurement time to start the measurement. Figure 3.6 is a schematic diagram of the experimental steps and current changes, including mechanical breaking of the nano-wire, formation of a nano-gap, and current generation when a molecule passes through the nano-gap.

During the measurement, the gap distance must be kept below the set nanometer length. To keep this constant, the piezo-controlled pushing rod must be adjusted in real time and precisely controlled while measuring the current. The electrode-to-electrode distance is estimated using the current equation for direct tunneling current.

$$I = const \exp\left(-\frac{4\pi}{h}\sqrt{2mwl}\right) \qquad Equation\ 3.1$$

In this equation, h, n, w, and l are Planck's constant, the mass of the electron, the work function of Au, and the separation distance, respectively. In this study, the mass of the electron, m, is used as 9.1 ×10-31 kg, and the work function of Au(111), w, is used as 5.25 eV. In this study, the effective mass and work function of the Au(111) surface were used instead of the gold nano-gap for ease of calculation. Here, the work function of gold in aqueous solution may be lower than the actual one, and the widening of the inelastic gold gap immediately after the breaking of atomic bonds after the formation of the nano-gap is not considered, so the nano-gap distance applied in the measurement may be larger than the set value.

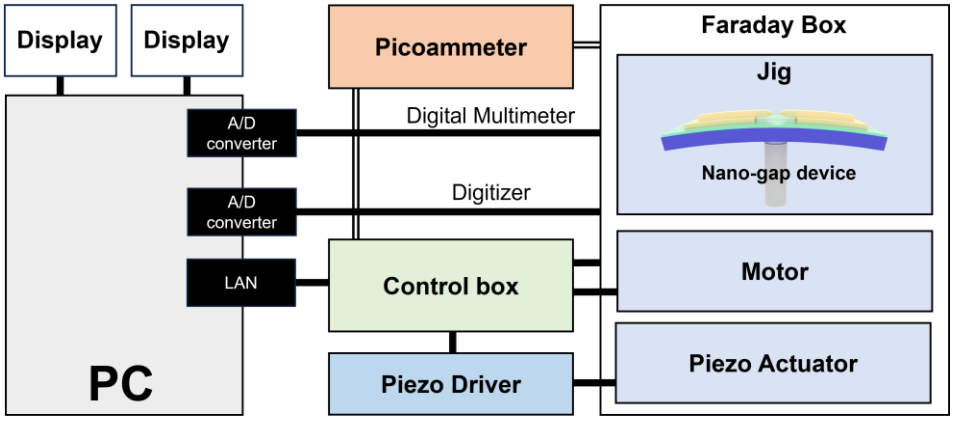

Figure 3.4 A simplified schematic representation of the MCBJ device system.

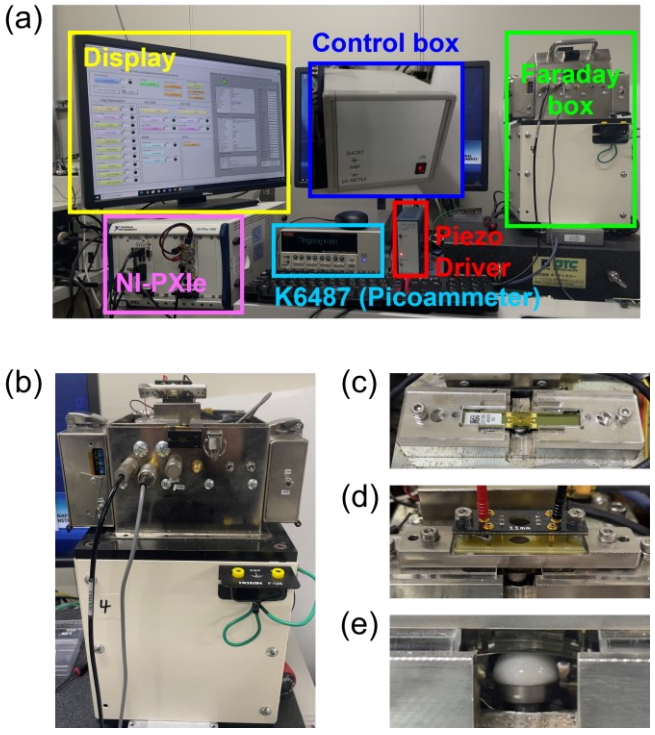

Figure 3.5 Images of the MCBJ measurement system. (a) Image of set up. (b) Faraday box. (b) MCBJ device chip inserted into the jig inside the Faraday box. (c) Image with cover on j ig. (c) Image showing the MCBJ device chip inside the jig bent due to the pushing r od rising.

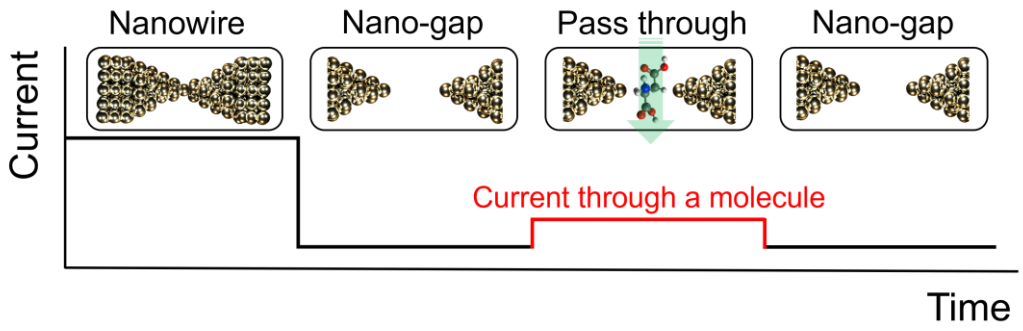

Figure 3.6 Schematic representation of the process of a single-molecule measurement. A sharp decrease in current when mechanically broken from a nano-wire state to a nano-gap. The tunneling current (red line) that occurs when a molecule passes through the nano-gap state.

3.3 Machine Learning Data Analysis

This research involves not only measuring single-molecules, but also discriminating molecules by analyzing individual signals from single-molecules with machine learning. The individual current signals are not labeled in advance and are extracted as features. After the single-molecule measurements described in Section 3.2, a baseline was estimated from raw data to extract individual current signals. In the process of extracting signals, the first step involves estimating the baseline level, which represents the underlying trend or average intensity of the signal. Following this, the noise level is estimated by calculating the standard deviation of data points within a specified region of the signal, capturing its inherent variability. The threshold of current intensity is set by multiplying the noise level by six times its value, creating a clear distinction between the desired signal and background noise. This process enables accurate identification and extraction of the target signal from the input data.

All extracted individual current signals were subjected to feature extraction for application in machine learning classification. The machine learning classifiers applied were random forest and XGBoost, as mentioned in Chapter 2, and the process described below is common to both. As shown in Figure 3.7, in this study, individual single-molecule current signals are extracted

by 13 factors consisting of maximum current (I_p), average current ($I_{avg.}$), duration (t_d), and a 10-dimensional normalized current factor, resulting in a 13-dimensional feature. Here, the 10-dimentional current factor is defined as the difference between the maximum current and the average current in each interval, dividing the individual signal into 10 parts along the time axis. This can be an indicator of the current variability in that bin. A larger difference between the two in each bin may indicate that the current in that bin is fluctuating unstably. This could indicate, for example, that a large current peak has occurred in the section during a certain time, or conversely that a number of current values below the average current level have been measured. This allows for a more detailed analysis of the current characteristics over the entire time period.

The features extracted in this way are used to train and test the machine learning classifier. Figure 3.8 shows the process of extracting features from individual signals and training and testing a machine learning classifier. To evaluate the performance of the trained model, we used K-fold Validation, which divides the data into multiple subsets (folds) and validates the model using each fold as a test set once. Figure 3.9 shows 10-fold validation, where the entire data was divided into 10 folds and evaluated. This method utilizes all of the data and allows for an objective assessment of the model's ability to generalize.

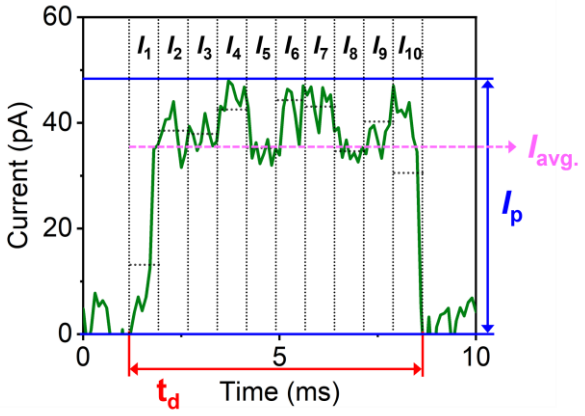

Figure 3.7 Single-molecule individual current pulse (Green solid line) and definitions of the features. The blue, red, and pink solid lines represent I_p, t_d, and $I_{avg.}$, respectively. The black dashed lines show the area of the current pulse divided into ten parts along the time axis.

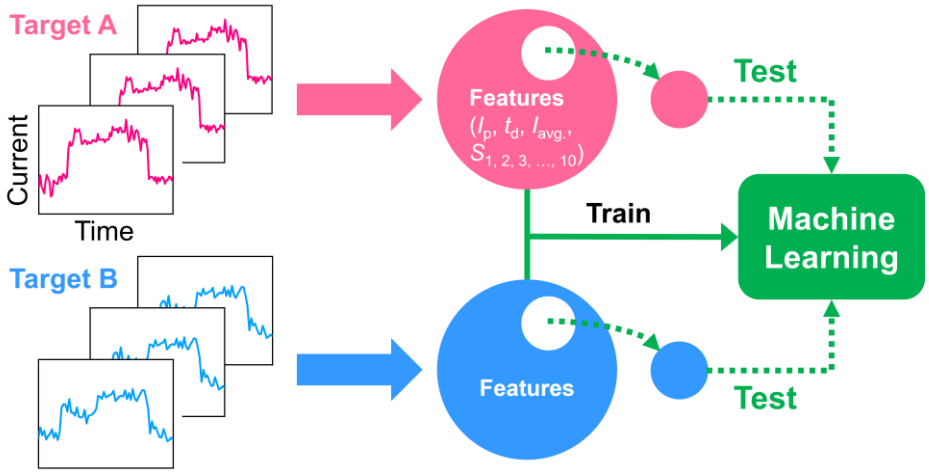

Figure 3.8 A flowchart of training a machine learning classifier through training and testing. Features includes factors such as peak current (I_p), duration (t_d), average current ($I_{avg.}$), and 10-dimensional normalized current for each pulse signal.

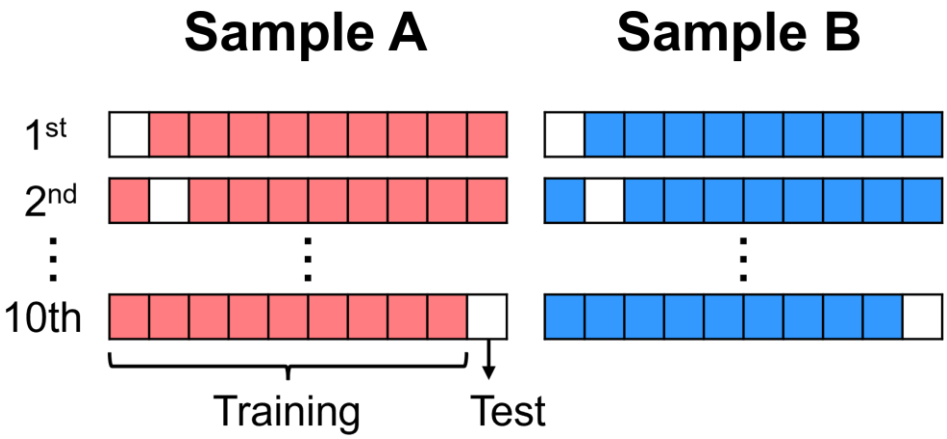

Figure 3.9 Schematic of 10-fold cross validation. 10-fold cross validation randomly divides the dataset into 10 equal-sized subsets. In each fold, one subset is held out as the test set, while the remaining 9 subsets are used for training. This process is repeated 10 times, with each subset used once for validation.

Chapter 4. Molecular Modification of Nano-gap Electrodes

This chapter discusses the first of several approaches to improving the discrimination accuracy of two molecules in single-molecule measurements where the discrimination accuracy is not high due to similarities in current characteristics. The introduction of machine learning to analyze single-molecule measurement data has improved molecular discrimination accuracy. However, it is not easy to discriminate molecules with high accuracy with current data from molecules that do not have discriminating factors. Therefore, this chapter discusses molecular modification approaches to improve molecular discrimination accuracy.

4.1 Introduction

Structurally similar molecules also have similar current properties, so molecules with similar structures measured in a bare nano-gap show similar current characteristics, which makes it difficult to improve single-molecule discrimination accuracy. A new approach to measurement is needed to discriminate these molecules with high accuracy. Therefore, I proposed a method to discriminate two molecules with similar molecular structures through molecular modification of nano-gap electrodes. As shown in Figure 4.1a,b, the two target molecules with similar structures to be discriminated in this study are Aspartic acid (Asp) and Leucine (Leu). Both target molecules belong to the amino acid group, which is one of the most important groups of molecules that make up the building blocks of life. Amino acids are used in the body for the biosynthesis of proteins, enzymes, peptides hormones, and some neurotransmitters. In addition to building biomaterials, amino acids are involved in many important metabolic processes in the body. In particular, there are 20 amino acids that make up proteins/peptides, and hundreds of metabolites. For this reason, the accurate discrimination of amino acids is important as a means to further understand the metabolic process of living organisms.

Techniques for analyzing amino acids have evolved steadily since Gordon, *et al*. first

separated and studied protein components using distributive chromatography in 1943.[66] Raman spectroscopy offers a number of analytical advantages, but unlike the constituent nucleobases of DNA, there is no way to amplify amino acids, making analysis difficult at low concentrations.[67–69] In the case of molecular imaging, the number of applicable molecules is limited because the probe must be prepared to match the target molecule.[70] T. Ohshiro, *et al*. measured the conductance of single molecules of amino acids passing through the nano-gap and analyzed their current characteristics for discrimination.[37] The current signals of the measured amino acid molecules were statistically analyzed using two characteristics, maximum current and duration, respectively. Figure 1.4, mentioned in Chapter 1, shows the amino acids classified from the I_p, t_d of several measured amino acid molecules. Some amino acids can be discriminated by both parameters, while others are difficult to discriminate due to the overlap of current characteristics. Despite the importance of amino acid analysis, the results of this study suggest that it is not easy to discriminate amino acids. It also suggests the importance of incorporating machine learning into molecular discrimination techniques.

As mentioned in Chapter 2, single-molecule conductance is related to the superposition of molecular orbitals. Therefore, chemical modification of nano-gap electrodes with appropriate molecules can optimize the molecular orbitals and thus improve the level of current. In a study to understand electron transport from one molecule to another, T. Nishino, *et al*. reported that hydrogen bonding by two -COOHs conducts electrons better than covalent σ bonds.[71] Based on these findings, I proposed to modify mercaptoacetic acid (MAA) molecules on nano-gap electrodes as a way to discriminate two target molecules with high accuracy, and to discriminate molecules that were not discriminated by conventional methods through machine learning-based time series analysis. As shown in Figure 4.1c, the MAA molecule has a -SH group and a -COOH group. The -SH group interacts strongly with the nano-gap electrodes made of Au.[72–78] If MAA is successfully bonded to the tips of both electrodes, these MAA-modified nano-gap electrodes are converted into electrodes with both tips composed of -COOH. Using this concept, it is expected to be possible to detect quantum tunneling enhancement through intermolecular hydrogen bonding interactions via -COOH groups.

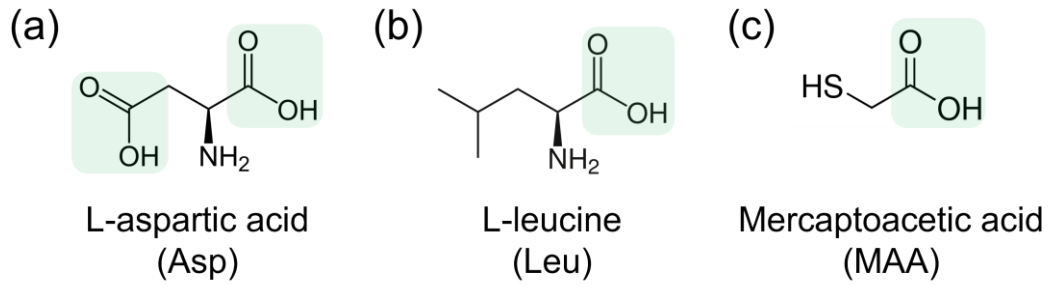

Figure 4.1 Molecular structures of the target molecule and the modified molecule. (a) L-aspartic acid. (b) L-leucine. (c) Mercaptoacetic acid. The light green caption shows the carboxy group of each molecule.

4.2 Methods

In this study, I compared the measurement results of two target molecules on bare gold nano-gap electrodes and the newly proposed molecularly modified nano-gap electrodes in this study. The two target molecules L-aspartic acid (Tokyo Chemical Industry Co., Ltd, Japan), L-leucine (Kanto chemical Co., Inc., Japan), and the modified molecule Mercaptoacetic acid (FUJIFILM Wako Pure Chemical Co., Japan) were used without any purification process using Milli-Q water as solvent. For the determination of pure target samples, an aqueous solution sample of 1 mM of each target chemical species was used. In this case, the pH of the pure solution of Asp and Leu at 1 mM concentration was 3.7 and 6.1, respectively. For all measurements where the nanogap electrode was modified with MAA, a 20 μM aqueous solution of MAA was used as the solvent. For the measurement of the Asp:Leu=3:1 mixture, a solution containing 750 μM of Asp and 250 μM of Leu was used, and for the measurement of the Asp:Leu=1:3 mixture, a solution containing 250 μM of Asp and 750 μM of Leu was used.

Each sample solution was prepared and used immediately prior to the measurement. A PDMS well about 5-6 mm thick was attached to the nano-gap device to contain the sample solution in the sensing part of the nano-gap device. Then, 15 μL of each sample solution was injected into the PDMS well, and Kapton tape was applied to prevent evaporation of the solution. The nano-gap device was inserted into the MCBJ measurement device and the nano-wire was mechanically broken to form a nano-gap as described in Chapter 3. The gap distances

of the nano-gap electrodes used for the measurements were set to 0.56 and 0.54 nm. Measurements were set to run for 5 minutes each at each gap distance and then repeated. Prior to all measurement steps, a blank solution was measured, excluding the target molecule. After completion of all measurement steps, the device was cleaned by injecting ethanol into the PDMS wells several times.

4.3 Results

4.3.1 Measurements with Bare Nano-Gap Electrodes

Pure Asp and Leu samples were measured using bare nano-gap electrodes. Before all measurements, the nano-gap device was filled with Milli-Q water as a blank measurement. Figure 4.2 illustrates the results of the blank measurement, presenting a current histogram and a current-duration 2D histogram. Both datasets indicate minimal current generation in Milli-Q water. Figure 4.3 depicts an idealized representation of two target molecules passing between gold nano-gap electrodes. However, in reality, they traverse the nano-gap electrodes in different orientations. As discussed in Chapter 3, the actual nano-gap distance is expected to be wider than the set value. Therefore, it is crucial to establish the appropriate nano-gap distance before the measurement. Starting from 0.66 nm, the approximate length of the target molecule, and gradually decreasing the setting by 0.02 nm, the frequency of signal generation by the molecule was considered. If the nano-gap distance becomes too small, the level of tunneling current between the gold nano-gap electrodes increases, and the single-molecule current is masked by noise. The goal of this study is to measure as many signals from different orientations of the molecules as possible and discriminate them through machine learning analysis. Therefore, the optimized nano-gap distance was set to obtain signals of sufficient quality for this analysis and to acquire many signals per unit time, with values set at 0.56 nm and 0.54 nm.

Figure 4.4 display selected three individual single-current pulses of Asp and Leu measured with bare nano-gap electrodes. The single-molecule signals generated by both target molecules exhibited akin current levels, and the overall shape of the signals did not possess distinctive features for discerning between the two molecules. The single-molecule currents measured in this study were numerous, and statistical analysis was employed to scrutinize the current

characteristics, given the diversity of currents originating from various molecular orientations. Figures 4.6a, b represent I_p histograms of single-molecule current signals for Asp and Leu, respectively. The current histograms for both target molecules are highly comparable. The I_p averages for Asp and Leu both showed a level of 19 pA, indicating that distinguishing between the two molecules based on I_p levels alone is not feasible. Figures 4.7a, b illustrate the current-duration 2D histogram. Both histograms suggest that distinguishing between the two molecules based solely on factors of current and duration is not clear. The measurement results of target molecules on bare nano-gap electrodes align with expectations, as the similarity in molecular structures did not exhibit sufficient differences in electron transport characteristics to distinguish between the current characteristics of the two molecules.

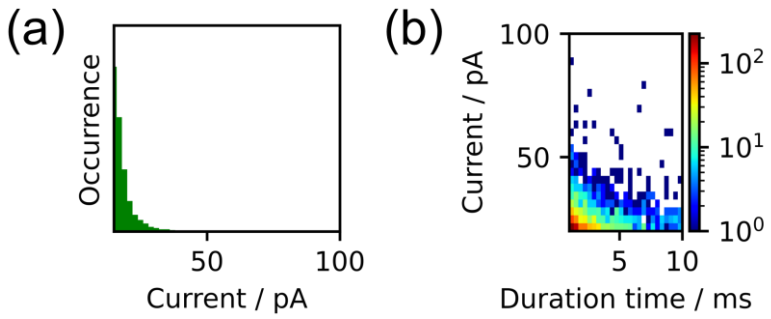

Figure 4.2 Measurement results in blank solutions(Milli-Q water). (a) Histogram of current measured from Milli-Q water. Frequency histogram of individual pulses according to the intensity of the current. (b) Histogram of current intensity by duration of current pulse.

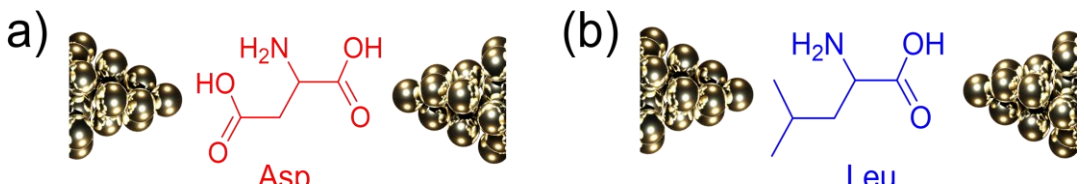

Figure 4.3 Measurement results in blank solutions(Milli-Q water). (a) Histogram of current measured from Milli-Q water. Frequency histogram of individual pulses according to the intensity of the current. (b) Histogram of current intensity by duration of current pulse.

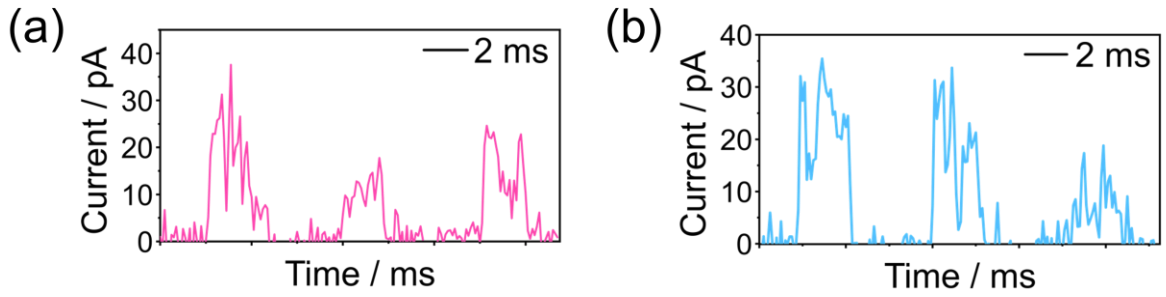

Figure 4.4 Individual current pulses measured in bare nano-gap electrodes. (a) Asp. (b) Leu. (a), (b) each represent three separate signals.

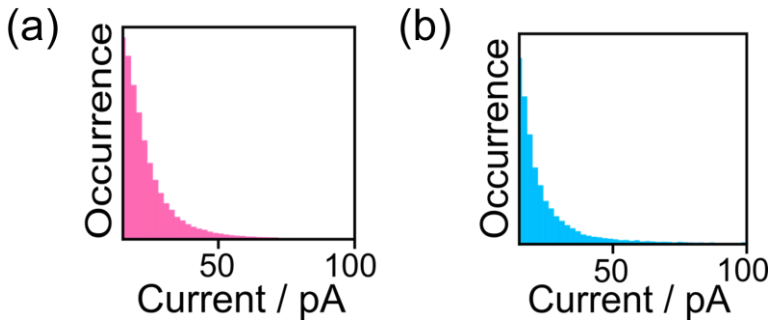

Figure 4.5 Histogram of the maximum current (I_p) measured in the MAA modified nano-gap. (a) Asp, (b) Leu.

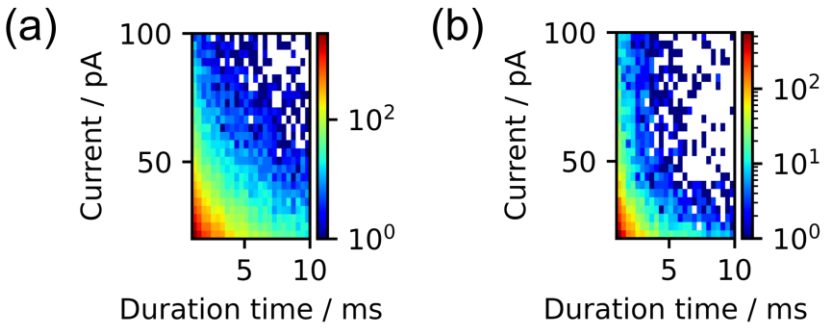

Figure 4.6 2-D Histograms of the current magnitude versus the duration of the current pulse. (a) Asp, (b) Leu.

4.3.2 Measurements with MAA modified Nano-Gap Electrodes

In this section, the results of measuring Asp and Leu molecules by modifying MAA molecules on nano-gap electrodes are discussed. As assumed in Section 4.1, the current level can be enhanced if the MAA-modified gold nano-gap electrodes are capable of hydrogen bonding interactions with the target molecules. MAA molecules have -SH groups that interact strongly with the Au electrodes and -COOH groups that can interact with target molecules via hydrogen bonding. Therefore, it was expected that Asp and Leu would interact with the -COOH group of MAA on both sides to different degrees when the molecule passes through the MAA modified nano-gap electrodes, as shown in Figure 4.7. Since the Asp molecule has -COOH groups on both terminals, it is expected to interact hydrogen bonding with both MAA modified nano-gap electrodes through the -COOH groups. The leu molecule, on the other hand, has a -COOH group on only one terminal, so it can only interact with one side. Therefore, for Asp, an enhancement of the current level due to hydrogen bonding interactions on both sides is expected.

Figure 4.8 shows the individual single-molecule signals of Asp and Leu measured on MAA-modified electrodes. Comparing these results to the case of bare nano-gap electrodes, the effect of MAA modification is clearly observed. In particular, the current pulses of Asp measured on MAA-modified nano-gap electrodes often exhibit a rectangular shape, unlike those in the other measurements. The I_p histogram in Figure 4.9 indicates that the current Asp level is higher than Leu. Compared to the measurements on the bare nano-gap electrodes, where the average of the I_p for both molecules was 19 nA, the average current for Asp increased to 23 pA and for Leu decreased to 16 pA.

As mentioned in Chapter 2, conductance of single-molecule depends on two competitive effects of tunneling distance and coupling strength between molecule and electrodes. The longer the tunneling distance offers the lower the single-molecule conductance.[79,80] Higher bonding energy leads to higher single-molecule conductance.[81] Asp molecule can form two stronger H-bond on both sides, whereas Leu can only form a stronger H-bonding via −COOH groups on either side. Additionally, te DFT calculations shown in Figure 4.10 explain that the interaction between -NH_2 and -COOH is weaker than the hydrogen bond between the two -COOH groups. Therefore, for Asp, interaction with the two -COOHs of both MAAs results in

current enhancement. This explains the stable current pulses shown in Figure 4.8a, where measurements on MAA-modified electrodes show that the current level of the single-molecule electrical signal of Asp is higher than that measured on bare nano-gap electrodes. The single-molecule current signals of Leu had relatively low current levels, but showed greater variability.

In addition to the interaction of MAA with Asp and Leu, we further investigated the effect of the dimer of MAA-MAA. Figure 4.11 shows the results of our investigation of MAA-MAA dimer formation. This is a measurement of a sample containing only MAA as a blank measurement for the measurement of the target molecule containing MAA. The three data in Figure 4.11 show that the measurement in MAA solution resulted in very few current pulses being observed. In principle, MAA-MAA dimers can form, but the incidence is very low due to the measurement conditions. The MCBJ instrument uses current feedback to control the gap distance. If a stable MAA dimer forms on the Au electrode, the base current is enhanced. Therefore, when MAA-MAA dimers are formed, the current feedback causes the electrode to contract. Therefore, even if dimers are formed, little current signal is observed.

The current pulses of Asp and Leu measured in the MAA modified nano-gap shown in Figures 4.8 and 4.9 show differences in the stability of the current signal and the current level. In the case of Asp, there are many current pulses with a rectangular shape, which indicates the stability of the current signal. To statistically analyze the signal stability of the measured current pulses, I normalized the current pulses in each measurement with I_p equal to 1. The four data in Figure 4.12 show 2D histograms of the normalized current-duration of all current pulses of Asp and Leu measured on bare nano-gap electrodes and MAA modified electrodes. The data to the right of each histogram shows the distribution of the normalized currents. Unlike the other three cases, only Asp measured at the MAA modified electrode shows a narrow distribution of current levels. The normalized average current values in the histogram are 0.84 and 0.80 for Asp and Leu, respectively, for the bare nano-gap. For MAA modified nano-gap electrodes, Asp and Leu are 0.90 and 0.82, respectively, with Asp showing the largest normalized current among the four measurements. The higher and narrower distribution of normalized currents indicates that the Asp molecule has a stable conductance with relatively little variability as it moves between the MAA modified nano-gap electrodes.

Amino acids are known to exhibit a wide range of single-molecule conductance due to

ionization in aqueous solution, and the I_p histograms show the variability of these measured currents well[38,82]. Here, I assumed that the hydrogen bonding between Asp and MAA via the -COOH group imparted a well-defined structure to the amino acid, which contributed to the fluctuations in the current levels of the Asp molecule measured at the MAA-modified nano-gap electrodes. It is important to note here that this study was conducted in aqueous solution for ease and potential use in molecular detection. In general, H_2O, a polar protonated solvent, inhibits hydrogen bonding. For example, CH_3COOH is known to form dimers by hydrogen bonding in the gas phase and in crystals, but not in aqueous solution.[83,84] In previous studies, it has been reported that the single-molecule conductance through DNA base pairs in H_2O is less than that of other non-protonated solvents[85], but in this study, single-molecule currents occur only when the target is positioned on the electrode or when the modified molecule is positioned on the electrode. Thus, the target molecule interacts with the electrode even when its structure is not in its most stable state. Other previous study has reported hydrogen bonding in aqueous solution environments.[86] In this study, the -COOH groups were shown to interact with each other and form transient H-bonds when Asp approached the -COOH groups of MAA as it passed through the gap of MAA-modified nano-gap electrodes. These results suggest that, unlike in the case of ensemble measurements, it is possible to observe localized minimum energy states that single-molecule measurements cannot reliably capture.

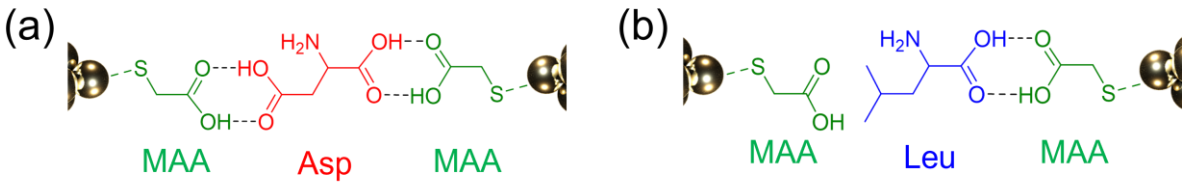

Figure 4.7 Schematic representation of Asp and Leu as they pass through an MAA-modified nano-gap. (a) Asp interacts with both MAAs via -COOH on both sides. (b) Leu interacts with the MAA with only one -COOH on each side.

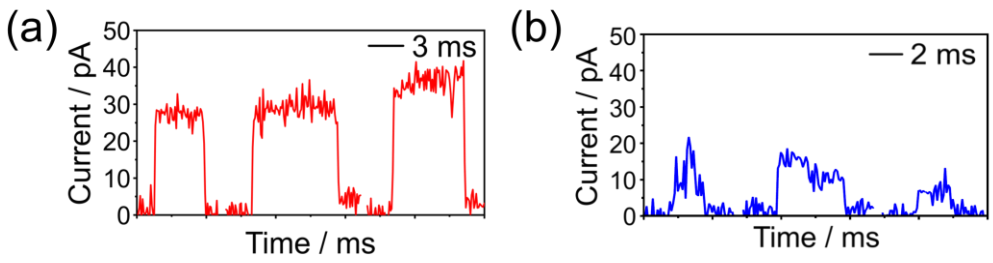

Figure 4.8 Individual current pulses measured in the MAA modified nano-gap. (a) Asp, (b) Leu.

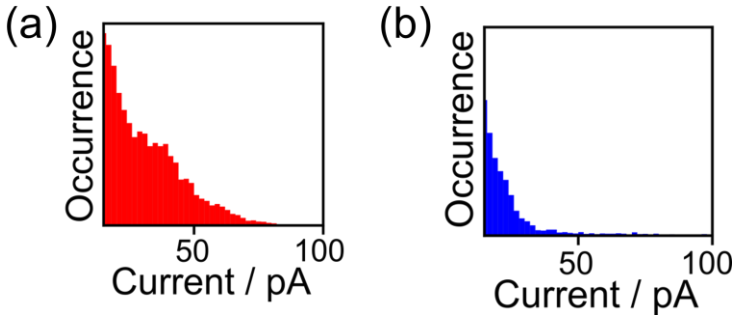

Figure 4.9 Histogram of the maximum current (I_p) measured in the MAA modified nano-gap. (a) Asp, (b) Leu.

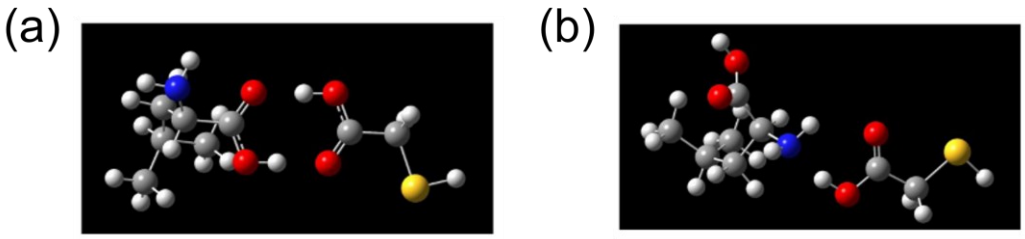

Figure 4.10 Schematic of calculated Leu and MAA hydrogen bonding (a) -COOH in Leu and -COOH in MAA. (b) -NH_2 in Leu and -COOH in MAA. DFT calculation method is B3LYP/6-31G(d,p) . Gaussian 09 were used. Optimization provided COOH-COOH bonding (a) and NH_2-COOH structures (b) as shown in Fig. S6. The bonding energy is 83 kJ/mol and 64 kJ/mol for (a) and (b), respectively.

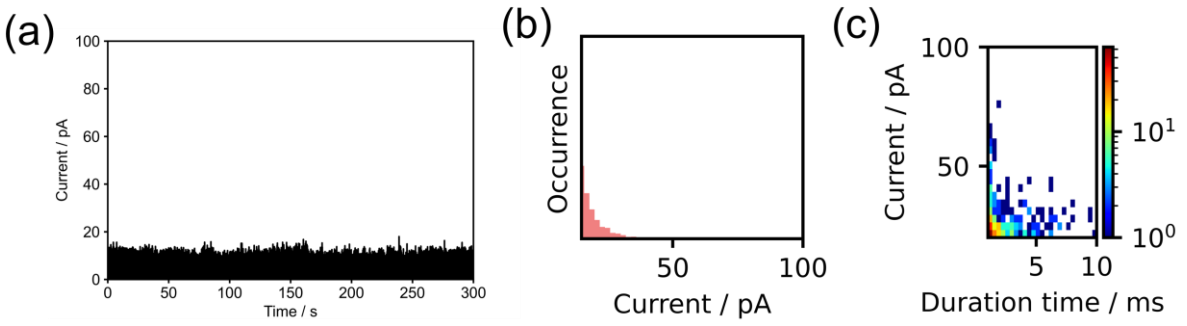

Figure 4.11 Results in MAA solution measurement. (a) Current profile of MCBJ measurement in MAA solution. (b), (c) Histograms analyzing the currents measured from 20 μM MAA solution. (a) Frequency histogram of individual pulses according to the intensity of the current. (b) Histogram of current intensity by duration of current pulse.

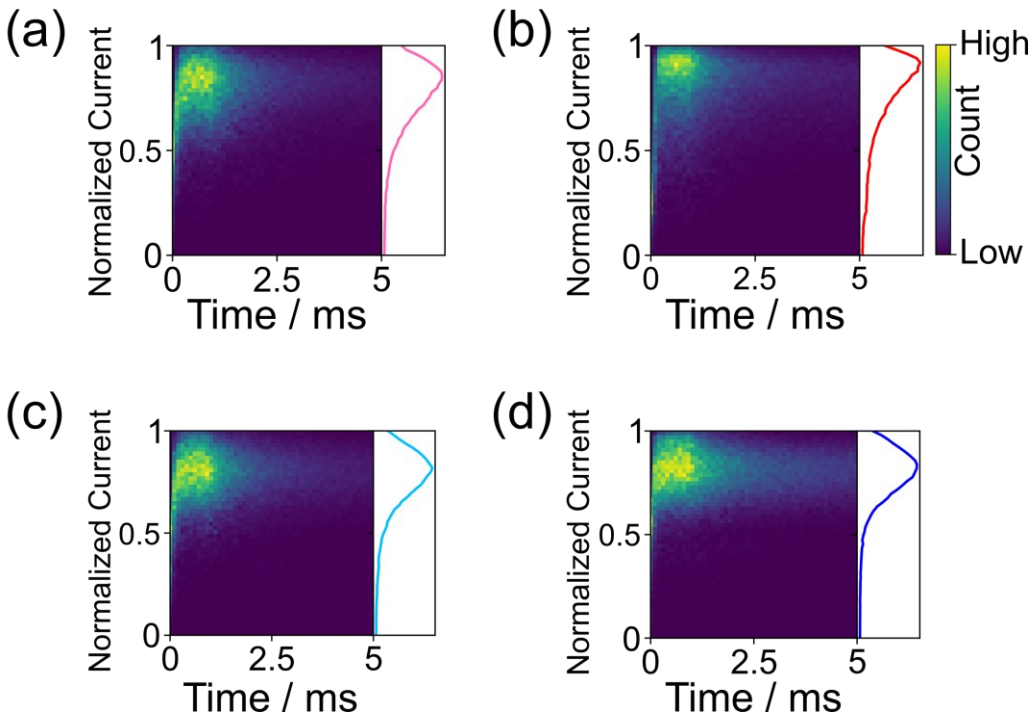

Figure 4.12 Histograms of normalized current and time for Asp and Leu at bare Au and MAA-modified nano-gap. The histograms present the results for (a) Asp at bare electrode, (b) Asp at MAA-modified electrode, (c) Leu at bare electrode, and (d) Leu at MAA-modified electrode. The right-hand side of each histogram shows a histogram of normalized current intensity.

4.3.3 Molecular Discrimination with Machine Learning

In the previous section, I measured Asp and Leu in two environments: bare nano-gap and MAA modified nano-gap electrodes, and observed and analyzed the current characteristics of each. Unlike the case of bare nano-gap, Asp and Leu measured in MAA modified nano-gap electrodes showed differences in current characteristics. As mentioned in Chapter 3, the measured data is used to train machine learning and discriminate molecules. The trained classifier predicts the proportions of the mixture from the measured data of a sample of the mixture.

In this study, the XGBoost machine learning classifier was used for molecular discrimination to classify the single-molecule currents of pure Asp and Leu samples measured on bare nano-gap electrodes and MAA modified nano-gap electrodes. Figure 4.13 shows the process of training the machine learning classifier. Here, 2,680 current pulses of each Asp,Leu measured on bare nano-gap electrodes and 1,180 current pulses each on MAA modified nano-gap electrodes were trained and tested by the classifier. As described in Chapter 3, the features used were maximum current (I_p), average current ($I_{avg.}$), duration (t_d), and 10-D normalized current factor, and training and testing were performed with 10-fold validation.

Figure 4.14 is a confusion matrix showing the results of machine learning classification of single-molecule current data of pure Asp and Leu samples measured on bare nano-gap electrodes. The prediction accuracies for Asp and Leu were 0.59 and 0.51, respectively. This result shows that machine learning discrimination of Asp and Leu is not accurate because their signal characteristics are similar. Therefore, the measurement method using a bare nano-gap electrode is not a suitable method for discriminating Asp and Leu. Figure 4.15 shows the machine learning classification results of Asp, Leu measured on MAA modified nano-gap electrodes. For MAA modified nano-gap electrodes, the accuracy of Asp and Leu was 0.80 and 0.78, respectively. As Figure 4.16 shows, if the discrimination accuracy using one signal is 0.77, the accuracy of statistical analysis using 13 signals is more than 99%. The modification of the nano-gap electrodes by MAA resulted in high accuracy classification of the single-molecule signals of Asp and Leu, which could not be distinguished by the measurement of bare nano-gap electrodes.

To verify that the method is effective in practice, the current signals obtained from single-molecule measurements of mixed solutions of Asp and Leu were analyzed to predict the mixing ratio. Two mixture solutions were used for validation: Asp:Leu=3:1 (750 uM: 250 uM) and Asp:Leu=1:3 (250 uM: 750 uM). The machine learning classifier was trained with the measured results of the mixture solutions on both bare and MAA modified nano-gap electrodes. After training, the single-molecule currents of the mixed solutions were discriminated individually, as shown in Figure 4.17. The mixing ratio is expressed as the fraction of the total signal that is counted as the predicted molecule. Figure 4.18 shows the predicted mixing ratio of a mixed solution measured on bare nano-gap electrodes based on training results from pure target samples measured on bare nano-gap electrodes. For Asp:Leu=3:1, Asp:Leu=0.52:0.48 was predicted, and for Asp:Leu=1:3, Asp:Leu=0.49:0.51 was predicted.

For the bare nano-gap electrodes, the two amino acids were not correctly classified in the mixed solution, as was the signal classification for the pure target solution. On the other hand, as shown in Figure 4.22, predicting the mixture ratio from the measured data on the MAA modified nano-gap electrodes predicted a ratio of 0.75:0.25 for Asp:Leu=3:1 and 0.23:0.77 for Asp:Leu=1:3. Figure 4.23 shows the current profile of Asp:Leu=1:3 measured on MAA modified nano-gap electrodes. The red and blue peaks represent the current pulses predicted by Asp and Leu, respectively, while the gray color indicates that neither is predicted. In summary, these results show that in single-molecule measurements using nano-gap electrodes modified with MAA molecules, hydrogen bonding can affect the single-molecule conductance properties.

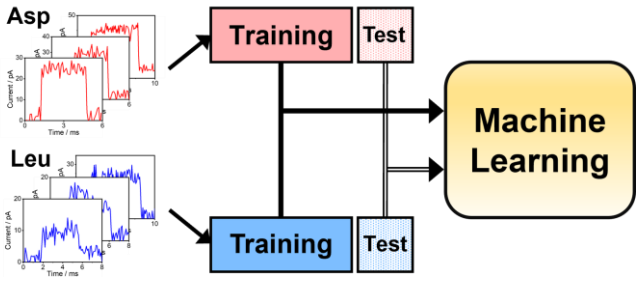

Figure 4.13 Training and prediction process for machine learning classifier using single-molecule individual current pulses.

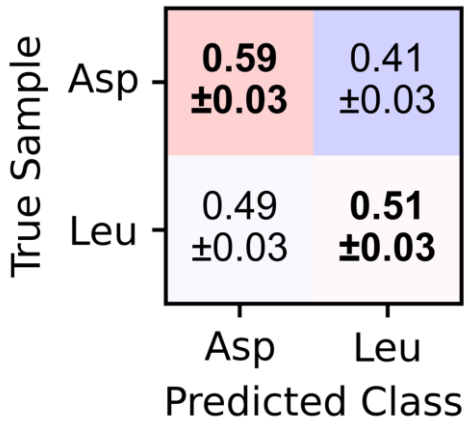

Figure 4.14 Confusion matrix of molecular prediction from a machine learning classifier trained on Asp and Leu single molecule current signals measured at bare nano-gap electrode. The F-measure is 0.55.

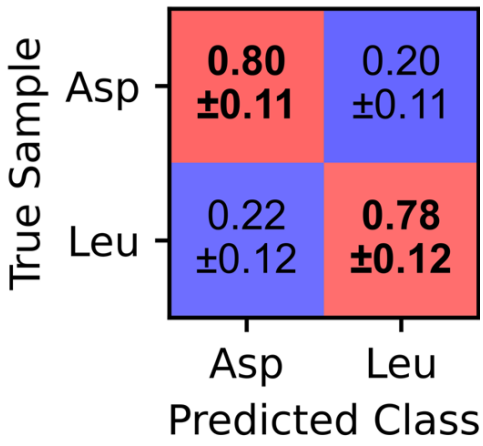

Figure 4.15 Confusion matrix of Asp and Leu predictions measured with MAA-modified nano-gap. The F-measure is 0.79.

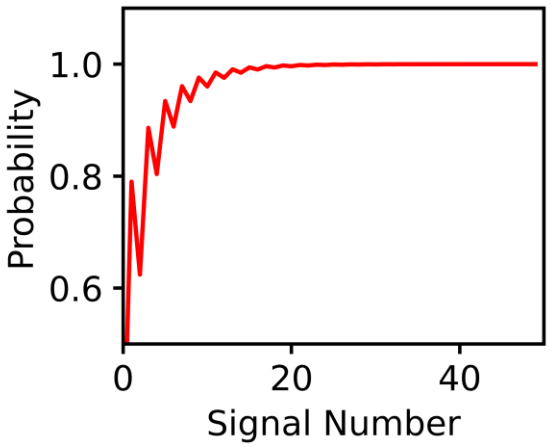

Figure 4.16 Correlation between the accurate predictability by majority vote and the number of signals.

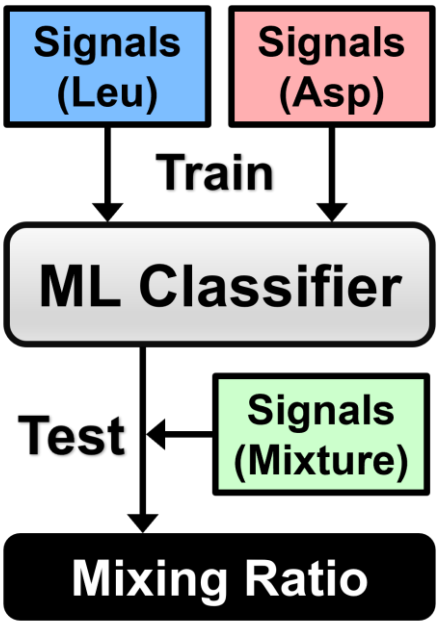

Figure 4.17 Flowchart of machine learning analysis for mixing-ratio prediction.

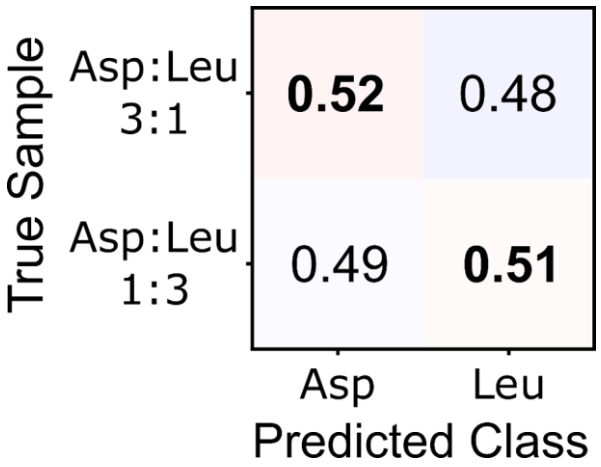

Figure 4.18 The predicted mixing ratios of Asp and Leu mixtures measured with bare nano-gap electrode. The upper row shows Asp:Leu=3:1 and the lower row shows Asp:Leu=1:3.

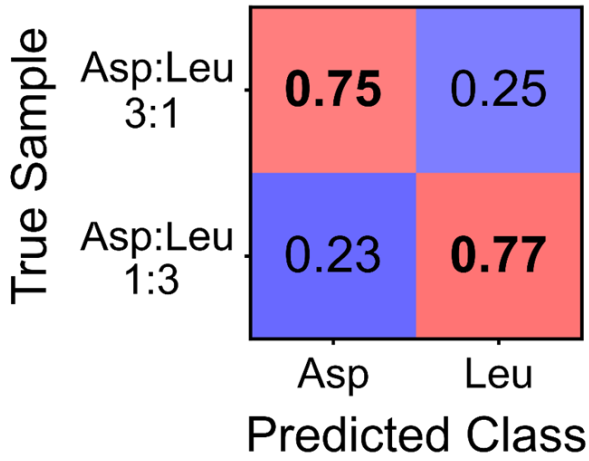

Figure 4.19 Prediction result of the mixing ratio of solutions from the ratio of mixtures Asp:Leu =1:3 (upper) and Asp:Leu = 1:3 (bottom). (c) Current profile measured in Asp:Leu=1:3 solution. The red, blue and gray signals represent Asp, Leu and invalid signals respectively.

4.5 Conclusion

For The results of this study demonstrate that the convergence of measurement methods to modify molecules suitable for nano-gap electrodes and machine learning-based data analysis is an effective way to discriminate between molecules with similar structures. Asp with two -COOH groups on both terminals showed a stable and high level of single-molecule current due to the hydrogen bonding interaction with MAA bound to the nano-gap electrodes. On the other hand, Leu, which has a -COOH group on only one terminal, can only hydrogen bond with MAA on one side. For this reason, Leu could not fully interact with both MAA molecules, which eventually led to an increase in the tunneling distance, which was lower than that measured in bare nano-gap electrodes.

This study demonstrated that differences in molecular behavior under certain conditions can be exploited to improve the discrimination accuracy of molecules. Interestingly, the interaction effect between the -COOH groups of MAA and Asp was observed even in aqueous solutions, where hydrogen bonding is typically known to be disrupted. This suggests that localized/transient interactions due to hydrogen bonding between molecules in aqueous solution can be monitored by single-molecule measurements. Subtle changes in single-molecule current behavior were detected and used to distinguish between two amino acids that were barely distinguishable by conventional methods and machine learning analysis. This concept is expected to enable the analysis of a wide variety of molecules in the future and is an important concept for the development of protein/peptide sequencing technology based on single-molecule measurements.

Chapter 5. Molecular Discrimination through Chemical Environment Changes

In Chapter 4, a novel approach to improving molecular discrimination accuracy by modifying nano-gap electrodes with appropriate molecules was discussed. Although this approach improved the accuracy of molecular discrimination, it can only be applied when there is a clear chemical difference between the target molecules, and appropriate modification molecules are required. Therefore, the application of molecular modification methods may be limited. Considering this limitation, I focused on the difference in behavioral changes of molecules due to changes in chemical environment. Molecules show behavioral changes due to changes in chemical environment. However, under the same chemical environment change, molecules will exhibit different behavioral changes. In this chapter, I propose an approach to molecular discrimination through differences in the behavior of single-molecules. This approach will also provide an opportunity to explore how changes in the chemical environment affect the behavior of molecules.

5.1 Introduction

One of the factors that determines the current properties of a molecule is its electronic structure, which is derived from its molecular structure. The structure of a molecule can be changed by changes in the chemical environment. For example, protonation/deprotonation of functional groups or intramolecular reactions due to changes in pH can cause changes in molecular behavior. Changes in the current properties of molecules due to protonation/deprotonation of functional groups by changes in pH have been reported by previous studies.

F. Chen, *et al*. observed differences in the behavior of chemical linker groups, -NH_2 and -COOH groups, due to changes in pH condition.[81] Figure 5.1 shows the conductance histograms

of diamine butane and dicarboxylic acid butane molecules measured at different pH environment, respectively. Diamine butane showed a characteristic conductance peak with the highest number of counts at pH 13, while the pH 10 condition showed fewer counts than pH 13. This is a result of the increasing proportion of $-NH_3^+$ as the pH decreases from 13 to 10. For pH 1, no characteristic peak was observed, which is interpreted to be due to the fact that the nitrogen atom of $-NH_3^+$, which is dominant at pH 1, does not have a lone pair and therefore cannot interact with the electrode. On the other hand, dicarboxylic acid butane showed a different behavior: dicarboxylic acid butane exhibited the highest characteristic conductance count at pH 13, and a lower level at pH 5. On the other hand, unlike diamine butane, dicarboxylic acid butane showed a peak at pH 1. This is because the oxygen atoms of -COOH have lone pairs, which allows it to interact with the electrode. Based on the results of this study, molecules with different types of functional groups are expected to show different behavioral changes in response to changes in the chemical environment.

A neurotransmitter, dopamine (DA) and its precursor L-DOPA were chosen as target molecules in this study. Neurotransmitters are essential for maintaining brain function and regulating emotions, and are strongly associated with many psychiatric and behavioral disorders. Detecting and monitoring neurotransmitters is therefore critical to understanding and treating many brain-related diseases. In particular, among DA-induced diseases, Parkinson's disease is the second most common psychiatric disorder after Alzheimer's disease, and is reported to affect more than 10 million people worldwide. The cause of Parkinson's disease is not yet clearly understood, but it is believed to be caused by a lack of DA in the brain.[87–89] DA is unable to cross the Blood-Brain Barrier (BBB) into the brain. Therefore, patients with Parkinson's disease are prescribed L-DOPA, a precursor of DA that can cross the BBB.

L-DOPA is a common treatment for Parkinson's disease because it effectively increases DA levels in the brain.[90,91] However, over time, the short half-life of L-DOPA leads to DA loss, which prevents sustained symptomatic relief. However, high doses of L-DOPA used to compensate for DA loss can lead to side effects such as dyskinesia, hypotension, and psychosis.[92] Therefore, it is important to continuously monitor L-DOPA and DA to ensure effective dosing of L-DOPA and assessment of the patient's condition.

Several analytical methods for simultaneous monitoring of L-DOPA and DA have been

developed in the various areas of electrophoresis[93–95], spectrophotometry[96–98], and electrochemical detection[99–108]. Electrochemical methods are the preferred platform for the detection of catecholamines, but distinguishing between the two molecules in solutions containing both DA and L-DOPA has been an issue in the field for decades. As the biosynthetic pathway in Figure 5.2 illustrates, the two molecules have similar chemical structures, which makes them difficult to discriminate due to their shared electrochemical oxidation potential.

Single-molecule measurements can detect small changes in the behavior of individual molecules. In addition, the application of machine learning to single-molecule current data analysis has greatly enhanced its potential for molecular discrimination.[42,45,52–59] In this study, I propose a novel approach to molecular discrimination using single-molecule measurements. The method exploits differences in the current behavior of molecules due to pH changes to induce changes in the current properties of the molecules. Machine learning techniques are then used to effectively discriminate between two molecules.

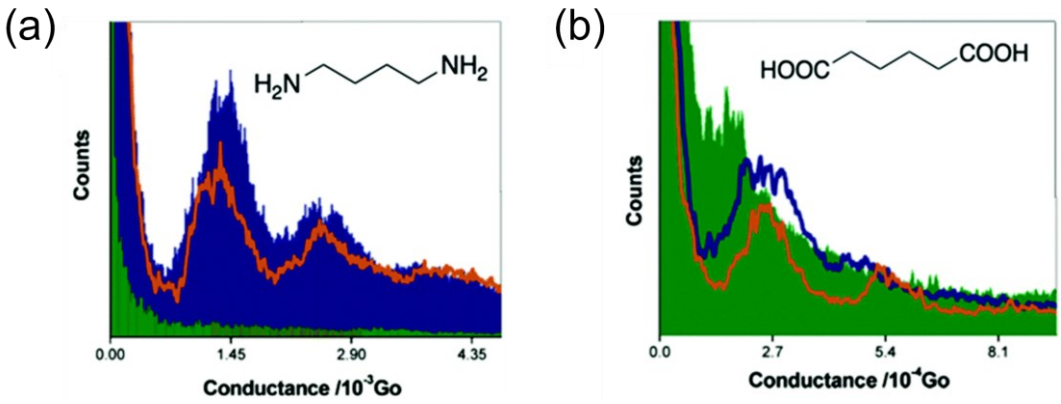

Figure 5.1 Conductance histograms of diamine butane and dicarboxylic-acid butane. (a) Diamine butane measured at pH 1 (green), 10 (red), and 13 (blue). (b) Dicarboxylic-acid butane measured at pH 1 (green), 5 (red), and 13 (blue). Figure adapted from "Effect of Anchoring Groups on Single-Molecule Conductance: Comparative Study of Thiol-, Amine-, and Carboxylic-Acid-Terminated Molecules" by F. Chen, *et al*., 2006, Article, *128*(49), 15874-15881, Copyright 2006 by ACS Publications, and used with permission.

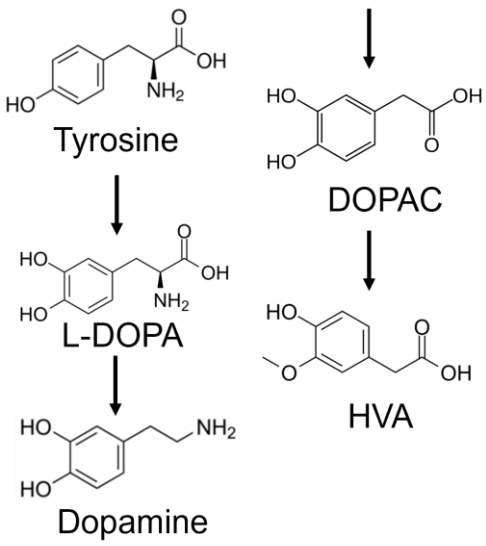

Figure 5.2 The metabolism pathway of dopamine (DA). L-DOPA is a precursor to dopamine.

5.2 Methods

The goal of this study is to measure two target molecules that are difficult to discriminate under neutral conditions and observe the difference in their behavior under acidic conditions. Therefore, the two target molecules were measured under different acidic conditions.

First, the MCBJ was used to measure each target molecule in neutral and acidic conditions. The target molecules L-DOPA (Sigma-Aldrich) and dopamine hydrochloride (Sigma-Aldrich) were used without purification. The two solvents used for the measurements were pH 7.6 Milli-Q water and pH 3, 0.001 M HCl aqueous solution. L-DOPA and DA were dissolved in both solvents at a concentration of 10 μM, so that a total of four solutions were used for each measurement.

After the measurement of the pure target solution was completed, two mixture solutions of different mixing ratios were prepared and measured. The two mixtures utilized solutions with the target solution dissolved in aqueous solutions of pH 7.6 Milli-Q water and 0.001 M HCl in

the ratios of DA:L-DOPA=3:1 and DA:L-DOPA=1:3, respectively. Among the ratios of the mixture solutions, 1 means 250 μM and 3 means 750 μM, and four sets of solutions were used for the measurement of this mixture.

To prevent evaporation of the solvent during the measurement, nano-gap device equipped with a PDMS well in solution was used, as described in Chapter 3. The PDMS wells were filled with 12 μL of the sample solution, and Kapton tape was attached to the top of the PDMS wells to prevent evaporation. The nano-gap distances set in the measurement were 0.58, 0.56, and 0.54 nm, and the bias voltage was 100 mV. The same machine learning algorithm used in Chapter 4, XGBoost, was applied to analyze the measurement data.

5.3 Results

5.3.1 Measurements under Neutral Conditions

Figure 5.3 shows the maximum current (I_p) histograms and 2D histograms of current-duration for L-DOPA and DA molecules measured at pH 7.6. The histograms of the two molecules show similar current levels, and the 2D histograms also show no differences to discriminate between the two target molecules. For a numerical comparison of the current values of the two molecules, the average value of I_p was analyzed, which was 19.5 pA for L-DOPA and 19.1 pA for DA. The analyzed data shows that the current properties of the two molecules are similar at pH 7.6, which suggests the possibility of similar electronic structures.

After analyzing the current properties through histograms, the current data from the two molecules were discriminated using XGBoost. The confusion matrix in Figure 5.4 shows the discrimination results of the two molecules, and the machine learning classifier discriminated DA with 53% accuracy and L-DOPA with 56% accuracy. These accuracies mean that the two molecules are almost indistinguishable from each other. As described in Chapter 3, the machine learning classifier has a good ability to discriminate between different current properties because it analyzes the current pulses of individual molecules based on the factors of maximum current (I_p), average current ($I_{avg.}$), duration (t_d), and 10-D normalized current. Therefore, the low discrimination results suggest a significant similarity in the current properties of the two

molecules measured at pH 7.6.

In the next step, the mixture ratio was predicted from the measured data of the mixture solutions using the machine learning classifier from the previous step. The ratios for the two mixed solutions were DA:L-DOPA=3:1 and DA:L-DOPA=1:3, respectively. The individual current signals measured in each mixed solution are predicted by the classifier as one target molecule. Therefore, the total ratio of the individual mixture measurement data predicted by each target molecule represents the mixture ratio of the sample. Figure 5.5 shows the results of the mixture ratio prediction for a mixture solution. The predicted mixing ratio for each solution was 39:61 for DA:L-DOPA=3:1 and 51:39 for DA:L-DOPA=1:3. The predicted mixing ratio was significantly different from the actual mixing ratio of the solution.

As a result, the single-molecule discrimination results at pH 7.6 indicate that the current properties of the two molecules are very similar and that they are difficult to discriminate. This shows that pH 7.6 is not an appropriate condition for discriminating between the two molecules.

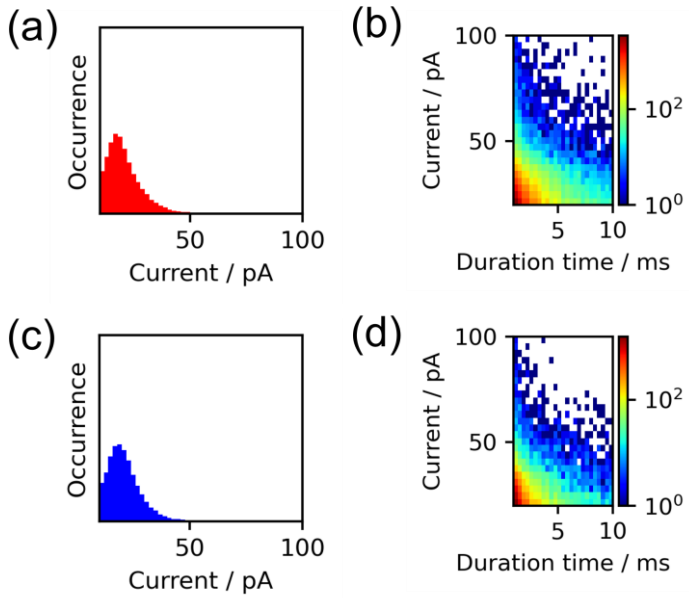

Figure 5.3 Histograms of maximum current (I_p) and 2D histograms of current-duration for two target molecules measured in pure solution at pH 7.6. (a), (b) show the case of L-DOPA, (c), (d) show the case of DA.

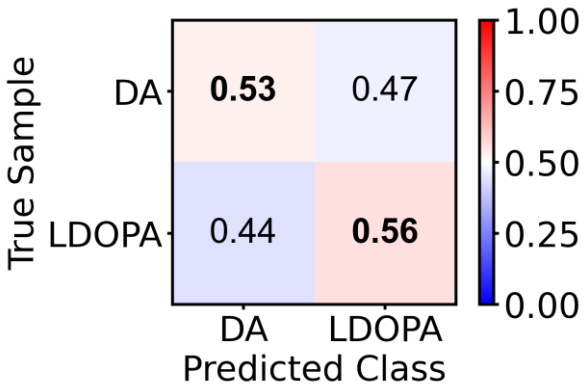

Figure 5.4 Discrimination results at pH 7.6 measured on pure solutions of L-DOPA (LD) and dopamine (DA).

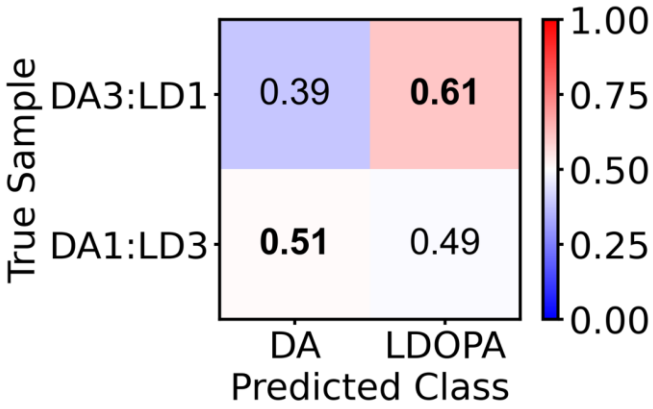

Figure 5.5 Mixing ratio prediction results for a mixture of L-DOPA (LD) and dopamine (DA) measured at pH 7.6.

5.3.2 Measurements under Acidic Conditions

In this section, the single-molecule measurements of DA and L-DOPA in acidic conditions are compared to the measurements in neutral conditions in the previous section. Figure 5.6 shows the maximum current (I_p) histograms and 2D histograms of current-duration for L-DOPA and DA molecules measured at pH 3. The histograms show that the current properties at pH 3 are different from those at pH 7.6. Unlike pH 7.6, the I_p histogram shows a clear

increase in current level for L-DOPA at pH 3. On the other hand, for DA, the current level seems to decrease at pH 3. Analyzing the average I_p levels at pH 3, we found 29 pA for L-DOPA and 16 pA for dopamine. Compared to the results at pH 7.6, there is an increase in current for L-DOPA and a decrease in current for DA. This suggests that the behavior of DA and L-DOPA is clearly different under acidic conditions. In particular, the large increase in current level for L-DOPA is noteworthy.

Figure 5.7 shows the discrimination results of the two molecules using XGBoost. DA and L-DOPA showed 81% and 86% accuracy, respectively, which is a significant improvement compared to the discrimination results at pH 7.6. The high accuracy of molecular discrimination at pH 3 indicates that the difference in current behavior between the two molecules is clear. From the mixed solution measurement data at pH 3 shown in Figure 5.8, the predicted mixing ratio was 82:18 for DA:L-DOPA=3:1 and 33:67 for DA:L-DOPA=1:3. Although the predicted mixing ratio is slightly different from the actual mixing ratio of the solution, it shows a high prediction accuracy compared to the case of pH 7.6. As a result, it indicates that the discrimination of DA and L-DOPA is more favorable in the pH 3 condition compared to the pH 7.6 condition.

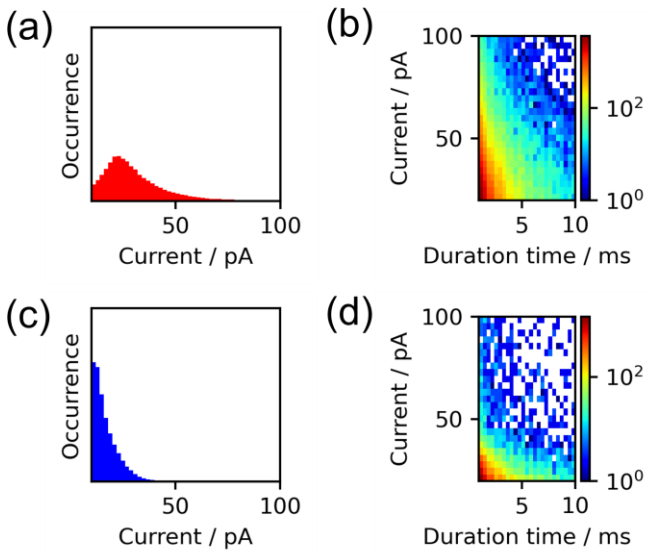

Figure 5.6 Histograms of maximum current (Ip) and 2D histograms of current-duration for two target molecules measured in pure solution at pH 3. (a), (b) show the case of L-DOPA, (c), (d) show the case of DA.

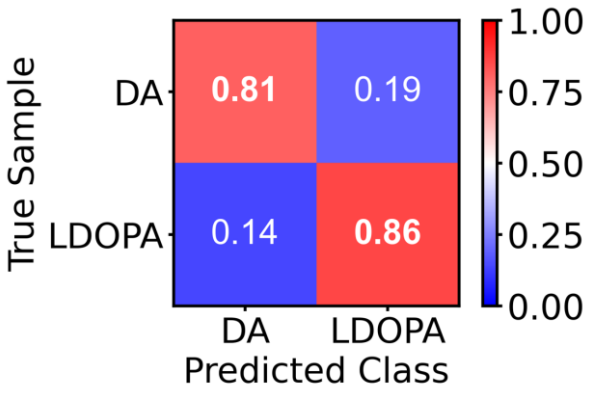

Figure 5.7 Discrimination results at pH 3 measured on pure solutions of L-DOPA (LD) and dopamine (DA).

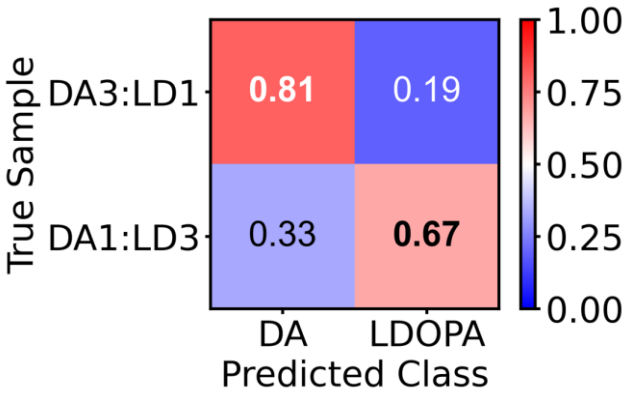

Figure 5.8 Mixing ratio prediction results for a mixture of L-DOPA (LD) and dopamine (DA) measured at pH 7.6.

5.6 Discussion of current behavior in Acidic Condition

The results in the previous section show that acidic conditions are more appropriate for discriminating between DA and L-DOPA than neutral conditions. At pH 3, the current level of L-DOPA increased, while the current level of dopamine decreased. From these results, it is expected that the change in pH affected the electronic structure of DA and L-DOPA. To further investigate the changes in the behavior of the molecules with changes in pH, we discriminated the same molecule at each pH condition. Figure 5.10a, c shows the I_p histograms of the same

molecule at each pH. The two histograms are an overlay of the I_p histograms presented in the previous section for each pH condition. Figure 5.10b, d shows the results of discriminating L-DOPA and DA measured at each pH from each other. The discrimination of L-DOPA shows 75% accuracy at pH 3 and 72% accuracy at pH 7.6. DA has 63% accuracy at pH 3 and 68% accuracy at pH 7.6. This discrimination result confirms that pH changes have a significant effect on the current behavior of DA and L-DOPA. Here, I inferred that L-DOPA has a higher discrimination accuracy than DA because its current behavior is more sensitive to pH changes. These results suggest that single-molecule measurements can detect changes in the behavior of molecules in response to changes in their chemical environment.

As a next step, I explored the factors that influence the change in the current behavior of both molecules in response to pH changes. One of the expected factors is the difference in the behavior of -COOH and $-NH_2$ under acidic conditions, which was introduced in Section 5.1. Except for the two -OHs on the benzene ring common to both molecules, DA has a $-NH_2$ group, while L-DOPA has both $-NH_2$ and -COOH. Therefore, as with diamine butane in Figure 5.1a, it is possible that the interaction of the molecule with the electrode was limited under acidic conditions due to the influence of $-NH_3^+$, resulting in a decrease in the current level. At pH 3, L-DOPA showed a significant increase in current. As shown in Figure 5.1b, the number of conductances generated by dicarboxylic acid butane at pH 1 is high. However, the level of conductance was lower at pH 1. Therefore, this result is not sufficient to explain the increase in current levels with L-DOPA.

Another factor in the expected change in current behavior is the intramolecular cyclization reaction of L-DOPA on the gold surface at pH 3. In a study to functionalize gold nanoparticles, N. Kalcec, *et al*. reported the behavior of L-DOPA and DA on the surface of gold nanoparticles at pH 3.[109] At pH 3, L-DOPA on the surface of gold nanoparticles undergoes intramolecular cyclization to form dopachrome. In contrast, DA stops at dopaminequinone and does not undergo intramolecular cyclization. Figure 5.11 shows a simplified reaction mechanism for L-DOPA and DA molecules at pH 3 proposed by N. Kalcec.

The nano-gap electrodes used for single-molecule measurements in this study are also made of gold, which means that they may have similar surface conditions to gold nanoparticles. I assumed that each reaction of DA and L-DOPA actually occurred on the surface of the Au nano-

gap electrodes, and investigated the changes in current behavior of the respective reactants and product molecules L-DOPA, DA, dopachrome, and dopaminquinone. According to the Breit-Wigner formula described in Chapter 2, the conductance increases as the HOMO level of the molecule approaches the Fermi level of the metal electrode.

To investigate this effect, the HOMO levels of the molecules were calculated and compared using Density Functional Theory (DFT). The DFT calculation method used is B3LYP/6-31G(d, p) utilizing Gaussian 09 software. Figure 5.12 shows the energy diagram of the Fermi level of Au and the calculated HOMO level of each molecule. According to the DFT calculations, the HOMO levels of L-DOPA and DA molecules were -8.60 eV and -8.53 eV, respectively. The calculated HOMO levels explain the similar current properties of the two target molecules under neutral conditions. Furthermore, the calculated HOMO levels of dopachrome and dopaminequinone were found to be significantly different. The calculated value was -6.19 eV for dopachrome and -9.49 eV for dopaminequinone. While dopachrome is closer to the Fermi level of gold, dopaminequinone is further away from the Fermi level of gold. The HOMO level calculations support the pH-dependent current properties of each molecule.

Further validation was performed to determine the current level based on the HOMO level according to the molecular structure. The molecule used for verification is hydroxyindole-2-carboxylic acid (HICA), which is one of the molecules similar in structure to dopachrome. DFT calculations of HICA showed a HOMO level of -5.62 eV. I measured HICA under neutral conditions and compared it to the current value of dopachrome. As Figure 5.12 shows, measuring HICA under neutral conditions resulted in an I_p of 26 pA. If we were to compare simply by HOMO level, the current level of HICA should be higher than that of dopachrome, but in reality, there are many factors other than HOMO level that determine the conductance of a single molecule. However, the fact that two similar molecules exhibit significantly higher current levels, as described by the Breit-Wigner formula, emphasizes the influence of the HOMO level of these molecules on the current level.

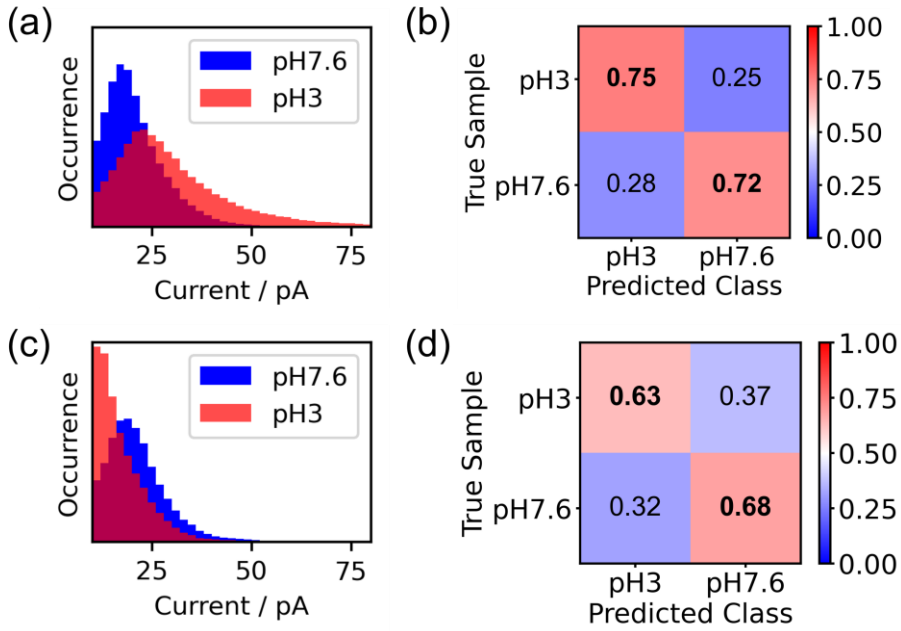

Figure 5.9 Mixing ratio prediction results for a mixture of L-DOPA (LD) and dopamine (DA) measured at pH 3. XGBoost was used as the classifier.

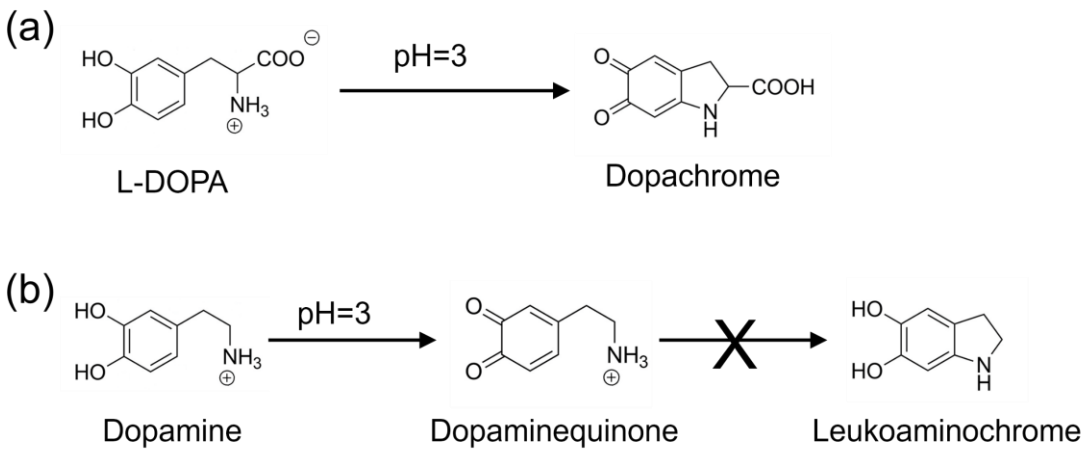

Figure 5.10 A simplified reaction mechanism for L-DOPA (a) and DA (b) molecules at pH 3 near gold structure proposed by N. Kalcec.

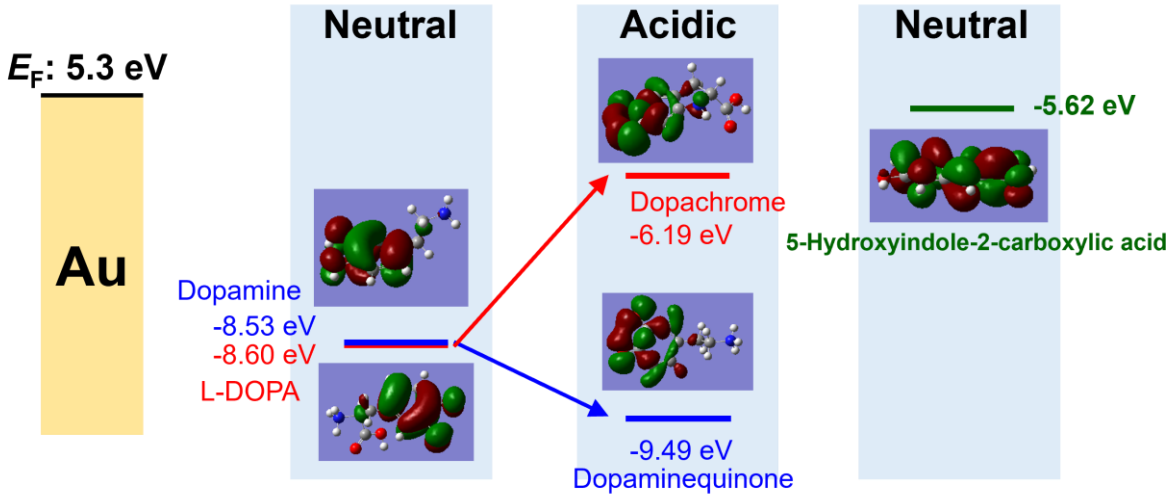

Figure 5.11 The results of HOMO level calculations.

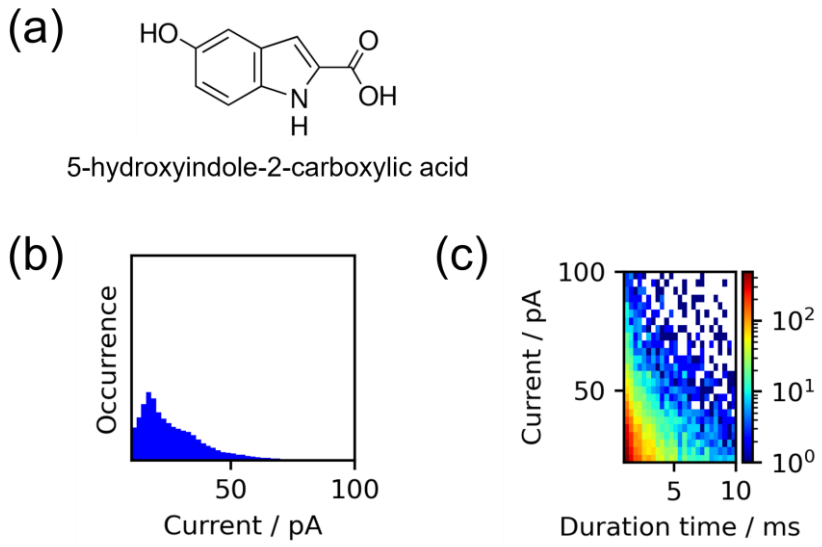

Figure 5.12 The results of 5-hydroxyindole-2-carboxylic acid. (a) The molecular structure of 5-hydroxyindole-2carboxylic acid. (b) I_p histogram of 5-hydroxyindole-2carboxylic acid. (c) 2D histogram of current and duration.

5.7 Conclusion

In this study, I proposed and validated an approach to differentiate molecules based on the differences in their behavioral responses to chemical environments. This methodology offers a simple yet effective means of distinguishing molecules with similar chemical structures and characteristics. Moreover, the target molecules of this research, DA and L-DOPA, hold various analytical and chemical implications. It is anticipated that this approach will prove beneficial for the analysis and discrimination of other significant molecules. Furthermore, this study demonstrates the ability to detect the impact of chemical environmental changes on molecular behavior through single-molecule measurements. This underscores the utility of single-molecule measurement techniques in exploring and comprehending the diversity of chemical events. These findings accentuate the significance of research and technological advancements at the single-molecule level across various domains.

Chapter 6. Direct Single-Molecule Discrimination without Pre-Training

Chapters 4 and 5 described strategies for distinguishing between two molecules with similar structures by modifying the nano-gap in the molecules or changing the chemical environment. The single-molecule measurement and machine learning data analysis methods described in the previous two chapters require classifiers to be trained with measurement data from pure target samples before analyzing real samples. However, it is difficult to apply this training process to all target molecules. In addition, due to the transformation of molecules by chemical equilibrium, accurate training data may not be accumulated, which can lead to errors in real sample analysis. Therefore, this chapter proposes a new algorithmic machine learning classifier that can immediately predict the composition of a sample without prior training. A training-free single-molecule discrimination method will enable single-molecule measurements to become a widely used chemical analysis tool.

6.1 Introduction

A method that can predict molecular composition by directly measuring samples containing multiple target molecules would minimize pretreatment and sample loss. In addition, the ability to analyze multiple targets simultaneously will significantly reduce time and cost. So far, the target molecules for single-molecule measurement and analysis are not limited to DNA/RNA[31–36], but have been expanded to a variety of biomolecules such as sugars[43], amino acids[110], proteins/peptides[37–41], and neurotransmitters[42]. It has been demonstrated that this technique can be applied not only to biomolecules but also to the analysis of various environmental molecules, such as explosives detection[111].

Many studies have measured a variety of single molecules, and the conductance of single molecules has a large variation. Therefore, statistical methods have been indispensable as a way to improve the accuracy of molecular discrimination, and conductance histograms have

been used as a traditional and common method. However, this only provides statistical information about single-molecule conductance. To overcome this limitation, machine learning analysis of single-molecule measurement data has been introduced, which has improved the discrimination accuracy of single-molecules.[42,45,52–59]

Conventional machine learning analysis methods require training data from pure sample solution measurements that contain only the molecule to be analyzed. However, considering the practical application of single-molecule measurements for specific molecule detection, it is not easy to prepare pure solutions for all expected impurity molecules in a sample. Rather, it may be relatively easier to prepare samples with varying concentrations of the target molecule in solutions containing various impurities. An example of such a method is the detection of a target molecule by manipulating the environment required for its release from a biological sample, either by promoting or inhibiting its release. This allows you to control the concentration or measurement signal of the target molecule and obtain accurate analysis results. Alternatively, a solution containing a reference molecule can be used to detect or measure the target molecule. However, even if it is possible to measure a solution containing only a specific target molecule, a machine learning classifier trained on that measurement data may not be applicable to real samples. This is because the measurement environment of the data used as training data may be different from the environment of the actual sample, i.e., environmental conditions, experimental conditions, presence of impurities, *etc*. may affect the measurement results, and these differences may limit the performance of the classifier.

For these reasons, the development of methods that can directly discriminate individual target molecules in mixed samples represents an important advance in the field of single-molecule measurement and discrimination. For example, such a method would allow for the discrimination of each target molecule in a mixed solution, without the need for a sample containing only a single species of target molecule. Therefore, this method has great potential for detecting and analyzing a variety of targets in complex environments such as practical biological samples, which will show promise for applications in a variety of fields. Therefore, the goal of this research is to develop a method for discriminating molecules that works with mixed solutions and does not require a single species target sample. Figure 6.1 graphically depicts the process of the conventional method and the new method. The substances targeted

in this study are deoxyguanosine monophosphate (dGMP) and thymidine monophosphate (dTMP), and Figure 6.2 shows the structures of the two molecules. Both dGMP and dTMP can be identified by single-molecule measurement and have been previously reported as target molecules in various studies.[31,46,112] In this study, the two target molecules were chosen in terms of discrimination of single-molecule signals by a novel machine learning model rather than in terms of the discriminability of each molecule in a mixture of the two molecules. In this study, I developed a novel analytical method to discriminate between the molecules utilizing single-molecule measurement data of only mixed solutions of dGMP and dTMP and compared it to conventional methods.

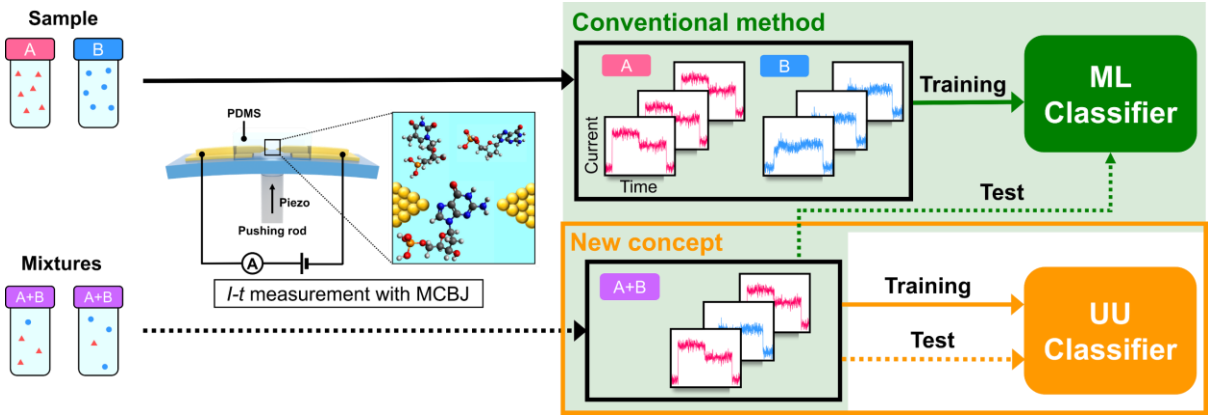

Figure 6.1 Flow chart of single-molecule classification. For single-molecule current measurements, the sample solutions were injected into a PDMS well, and the chips were bent with a finely controlled push bar with a piezoelectric device to form a nano-gap, after which the current was measured. The green box represents the conventional method, while the orange box represents the new concepts. The solid lines show the process for each individual sample, and the dashed lines show the process for the mixture.

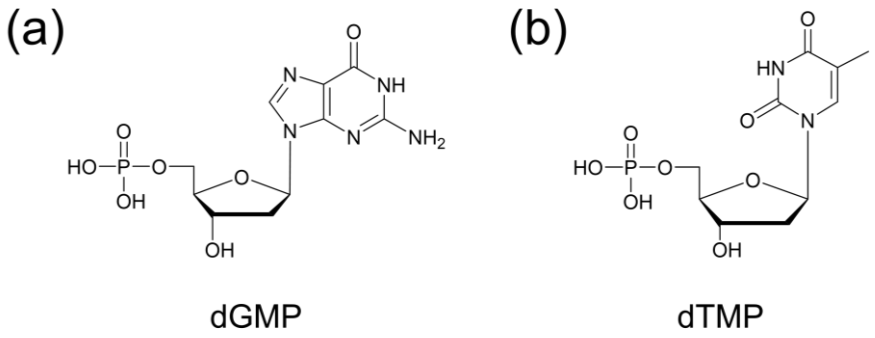

Figure 6.2 The molecular structures of dGMP and dTMP.

6.2 Methods

The goal of this study is to discriminate between the two target molecules using the conventional method and the new method, and to determine whether the new concept is effective in molecular discrimination. Therefore, dGMP and dTMP solutions were first prepared for the conventional method and measured by MCBJ. dGMP (Sigma-Aldrich) and dGTP (Sigma-Aldrich) were diluted in Milli-Q water without further purification. The concentration of each solution of dGMP and dTMP used in the measurement was 10 μM. After measuring pure target sample solutions, mixtures of the two species to be analyzed in both the conventional and new methods were measured. In this study, two types of mixture samples were used, and the dGMP, dTMP ratios in each sample were dGMP:dTMP=3:1 (750 μM dGMP and 250 μM dTMP) and dGMP:dTMP=1:3 (250 μM dGMP and 750 μM dTMP), respectively. PDMS wells were installed in the nano-gap device to contain the solutions. The nano-gap distances used for the measurements were 0.58, 0.56, and 0.54 nm.

After the single-molecule measurement was completed, machine learning was used to identify the two molecules. Each of the 830 pulse signals was trained and classified using supervised machine learning with a random forest (RF) classifier in scikit-learn version 0.24.2. For validation, a 10-fold CV was performed and the mean and standard deviation values provided the classification rate and error. In the mixed solution assay, the RF supervised machine learning classifier was trained with 1000 dGMP and dTMP signals each. Signals with Ip > 20 pA and t_d > 1 ms were analyzed. The signals in the mixture were classified one by one using

the trained classifiers. The analysis was performed using Python 3.10.4. The UUC and weighting KDE source code was written by hand using Python 3.10.4. The 1000 signals and features from the mixture are the same as in the traditional method. A Gaussian kernel is adopted, and the bandwidth is determined by Silverman's rule.

6.3 Results

6.3.1 Conventional Single-Molecule Discrimination

Conventional methods predict the mixing ratio of a mixture based on data trained with measurements from solutions containing only the target molecule. Therefore, the two target molecules were first measured in pure solutions containing only the target molecules. A tunneling current pulse occurs when a single molecule passes between nano-gap electrodes, as shown in Figure 6.3. Figure 6.4 shows a histogram of the maximum current (Ip). The average maximum current for dGMP and dTMP was 32 pA and 25 pA, respectively. As reported in previous studies, Figures 6.3, 6.4 show that dGMP has higher current levels than dTMP. The reason for the difference in conductance between the two molecules is that the HOMO level of dGMP is closer to the Fermi level of Au, the conduction pathway, than the HOMO level of dTMP. Although there is a difference in the average conductance between the two molecules, a histogram of the actual current values shows some overlap. In the histograms of both target molecules, a low current signal was observed at the 20 pA level. This low current signal is caused by the bridge structure between the nano-gap electrodes and the single-molecule charge transport, wherein electron transport through the lower molecular orbitals of the ribose sugar is responsible for this low current.[113] This is why we observed a large overlap in the maximum current histograms of the two molecules, and why histogram-based analysis methods that rely on Ip alone do not achieve accurate discrimination. Therefore, machine learning was introduced as an additional statistical analysis method.

In the conventional method, a machine learning classifier is trained using individual single-molecule current pulses obtained by measuring pure solutions for each single target molecule. Each solution is used as training data along with a molecular name (label). The classifier learns the current signal pattern based on the molecule name and uses it to discriminate the current

signal of the molecule in the separately measured mixed solution and to predict the mixing ratio. The discrimination results for the two molecules measured in pure solutions of each target were 0.75 for dGMP and 0.82 for dTMP, as shown in the Confusion Matrix in Figure 6.5. This approach demonstrates the ability of a machine learning classifier trained on data measured in a solution containing only a single chemical species to discriminate between molecules. To verify the classification ability of the machine learning classifier, we predicted the mixing ratios of the target molecules using the classifier trained on the current signals of each molecule in the previous step. Figure 6.6 shows the histograms of Ip measured in two mixtures with dGMP:dTMP=3:1 and dGMP:dTMP=1:3, respectively. Comparing the two histograms, the dGMP:dTMP=3:1 solution shows a higher conductance because it contains more dGMP, which has a relatively high conductance. Figure 6.7 shows the process of discriminating the mixtures using a machine learning classifier trained on each of the pure target solutions measured and analyzed in the previous step. The signals from the two mixture solutions were analyzed and discriminated to predict a mixture ratio of 64:36 for dGMP:dTMP=3:1 and 39:61 for dGMP:dTMP=1:3. As the discrimination results in Figure 6.7 show, the discrimination accuracy of each nucleotide target molecule varies individually, which may underestimate the predicted ratio of abundant nucleotides. In the next part, the same data as the mixture solution data used in this part will be used to predict the mixing ratio.

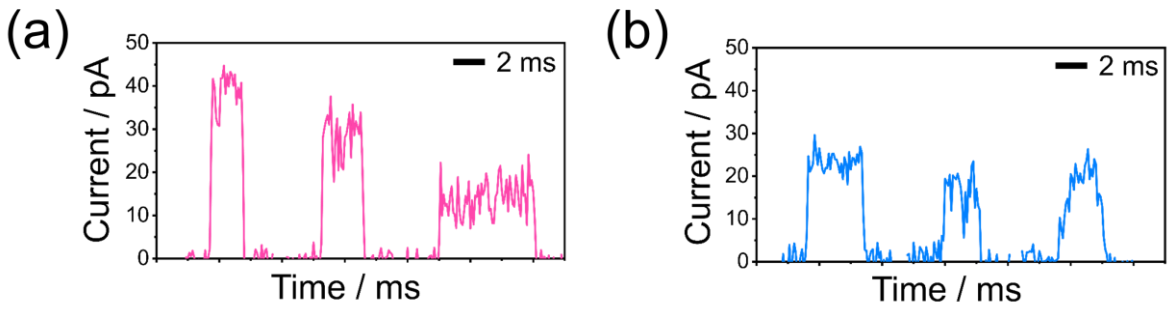

Figure 6.3 Individual current pulses of (a) dGMP and (b) dTMP.

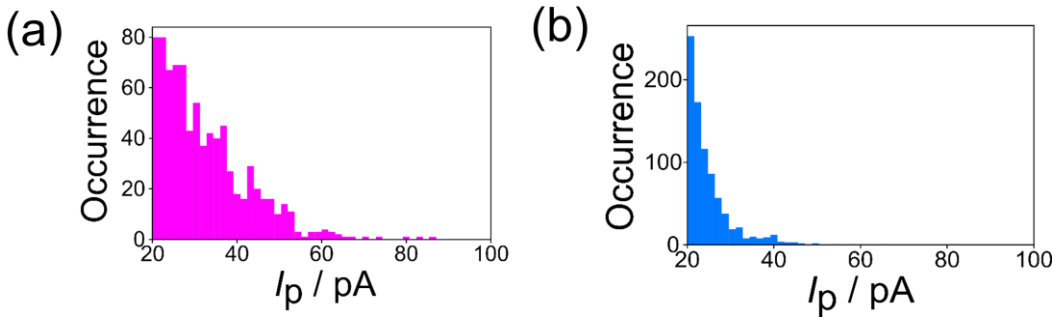

Figure 6.4 Histograms of the maximum current (I_p) for (a) dGMP and (b) dTMP.

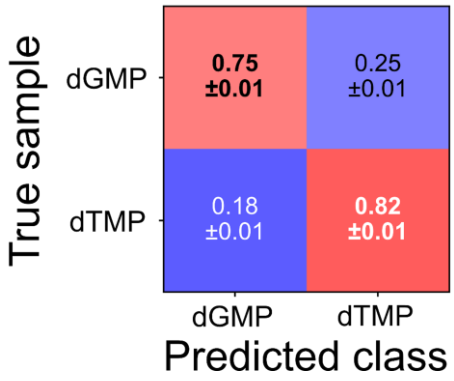

Figure 6.5 I_p histograms measured in two mixture solutions, dGMP:dTMP=3:1 and dGMP:dTMP=1:3.

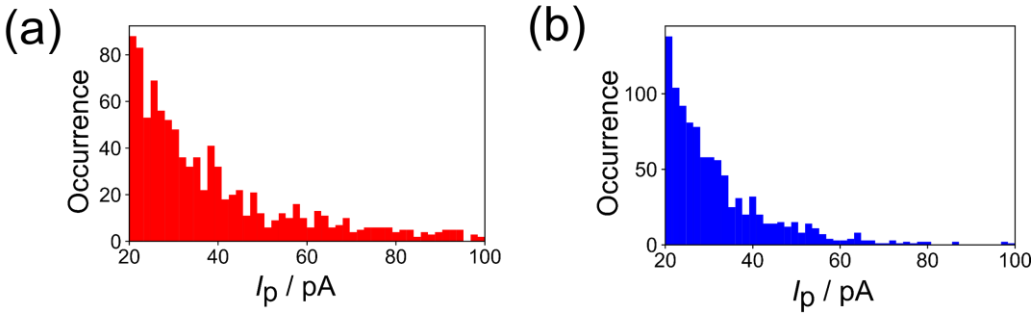

Figure 6.6 I_p histograms measured in two mixture solutions, dGMP:dTMP=3:1 and dGMP:dTMP=1:3,

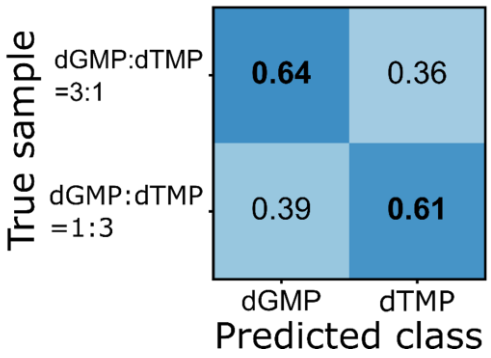

Figure 6.7 The results of predicting the mixing ratio of mixtures based on trained data.

6.3.2 New Concept of Single-Molecule Discrimination

In the previous section, we saw that when predicting the mixing ratio of a mixture, the more abundant of the two molecules in the mixture tends to be underestimated. This can limit the performance of the classifier because, as mentioned in the introduction, the measurement environment of the pure target sample may be different from that of the real sample. Therefore, as a way to minimize this effect, this part presents a method and results for predicting the mixing ratio of a mixed solution directly using only the mixture solution. Figure 6.8 illustrates the process of predicting a miscibility ratio from a mixture solution.

Given the data of a mixed solution of two molecules, the discriminative boundary of the two molecules must be determined in order to discriminate between the two unlabeled molecules. This was estimated directly from the data obtained from the two mixtures using Unlabeled data and Unlabeled data Classification (UUC) with Kernel Density Estimation (KDE).[114] Figure 6.9 shows a conceptual diagram of UUC, a method for determining discriminant boundaries from data from mixed solutions in which two classes (each molecule in this study) are present in different concentrations. In Figure 6.9, the red and blue colors represent mixtures with different mixing ratios, and the triangles and circles represent classes. Each mixture solution contains different concentrations of the two classes, and the classes are not known in advance. The main goal of UUC is to discriminate between the two classes based on which class is more abundant in this solution. Figure 6.10 illustrates the concept of KDE, which is a nonparametric statistical technique that estimates a probability density function directly from observed data.

Nonparametric statistics are methods that perform statistical tests by calculating probabilities directly from the given data, regardless of the characteristics of the population being sampled. So, intuitively, KDE calculates the probability density by adding a Gaussian kernel obtained from each observed data point, similar to how a histogram is generated by adding data points. The advantage of this method is that a smoother probability density distribution can be obtained with less data than a histogram. Therefore, by using KDE, a smoother representation of the probability distribution can be obtained, which allows for a more accurate distribution with less data. In this study, the Gaussian kernel was centered on directly measured data points. The UUC was determined by calibrating the probability density distributions of the two classes with KDE. However, since the proposed method is based on the principle of higher probability density in regions of higher concentration, it can also be applied when the concentration relationship of two unlabeled data mixtures is known.

To compare the conventional method with this new concept, a discrimination task was performed using the same features extracted from the same data as in the previous part. The UUC machine learning classifier was trained to predict the molecules using only the signals in the mixture, and Figure 6.11 shows the results of predicting the mixing ratio of the mixed solution. For dGMP and dTMP ratios of 3:1 and 1:3, the corresponding signals were predicted to be 76:24 and 22:78, respectively. Figure 6.12 shows the comparison of the discrimination results of the new concept with the results of the traditional method.

The electronic structure of the electrode affects the electrical transport properties of a single molecule, i.e., when a molecule is adsorbed on the surface of the electrode or the shape of the electrode changes, the electronic structure changes, and these changes can affect the signal of the single molecule.[115–118] In recent years, various machine learning methods have been developed, and these methods are used to analyze and model data. Unsupervised learning can be used to discriminate data without explicit labels, and it is used in conjunction with supervised learning. In other words, unsupervised learning is a learning method that does not require labels to discriminate patterns in the data. This unsupervised machine learning analysis was used to discriminate I-z traces in single-molecule measurement data.[54] However, as Figure 6.13 shows, traditional unsupervised machine learning methods cannot adequately discriminate data measured in two mixed solutions. The new UUC method applied in this study is able to

discriminate between the two molecules from data measuring only the mixture. In particular, the method avoids errors caused by changes in the environment and classifies single-molecules with higher accuracy than conventional methods.

In conclusion, this method has the potential to provide more reliable and accurate classification under a variety of conditions. Figure 6.14 shows the molecular prediction results obtained using UUC from the current profile of a dGMP:dTMP=3:1 mixed solution. The red and blue signals are the signals of dGMP and dTMP, respectively, predicted by UUC. This means that the signals of single-molecules from a mixture measurement can be individually differentiated. Figure 6.15 shows current profile images of measurements with each solution, including the blank solution (Milli-Q water). Figure 6.15 Current profiles of pure solution measurements of dGMP and dTMP, represented by a and b, show that dGMP has many current pulses with relatively higher current values than dTMP. For Figure 6.15 c, d, which shows the currents profiles in two mixture solutions of dGMP:dTMP=3:1 and dGMP:dTMP=1:3, it shows that there are many current pulses with higher values in solution c, which contains more dGMP. Figure 6.15e shows a measurement in Milli-Q water, showing that no current signal was generated. Figure 6.16 shows a duration histogram of the single-molecule current pulses from each solution.

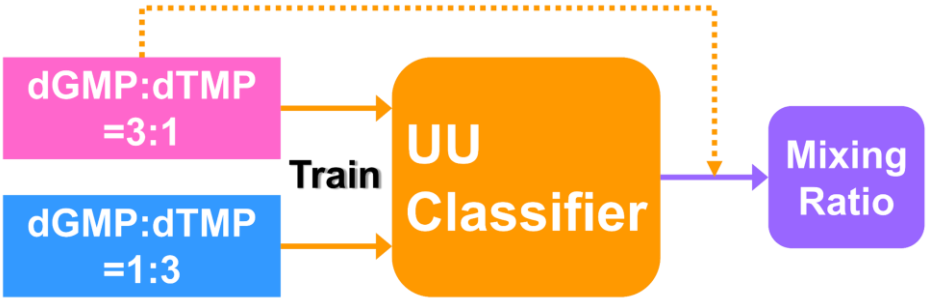

Figure 6.8 Process of training and identifying with data from mixtures only.

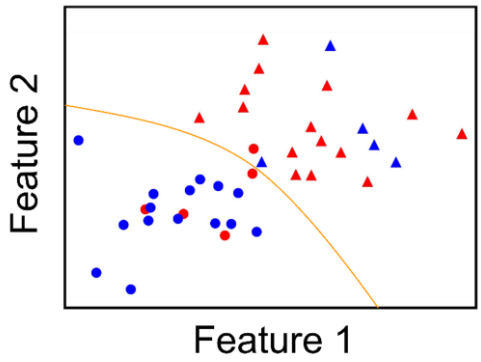

Figure 6.9 Schematic image of UUC. The red and blue colors represent two types of mixtures with different concentrations of the two classes. The circles and triangles represent each class. The UUC method determines the orange curve, which represents the boundary between two classes.

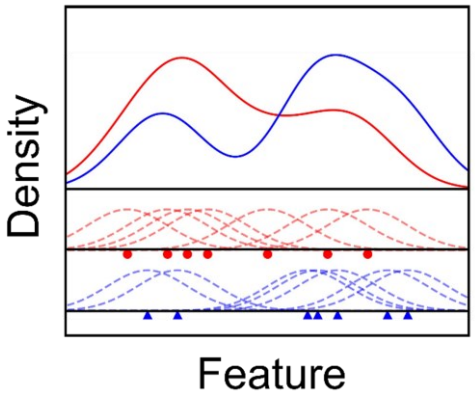

Figure 6.10 Schematic image of the KDE for estimating the probability density function in the feature space. The red and blue dots and dashed lines indicate the data points and their Gaussian kernel, respectively. The solid curves represent the sum of the dashed lines, which represents the kernel density estimate.

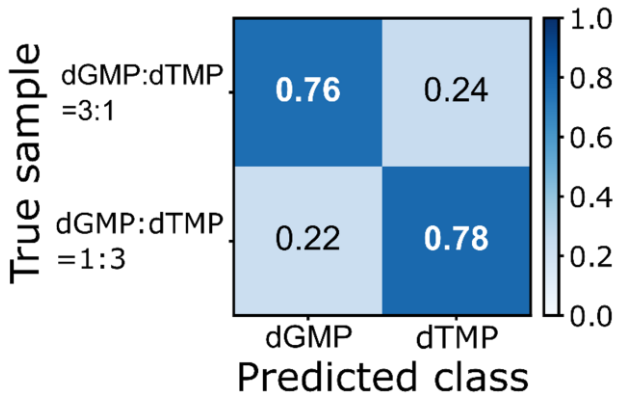

Figure 6.11 The result of predicting the mixing ratio of two mixtures with data trained on the mixture only.

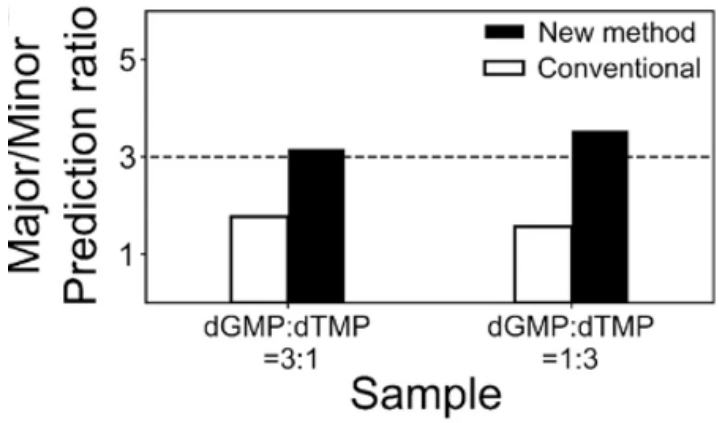

Figure 6.12 Comparison of the performance of the new and old methods with respect to the prediction ratio.

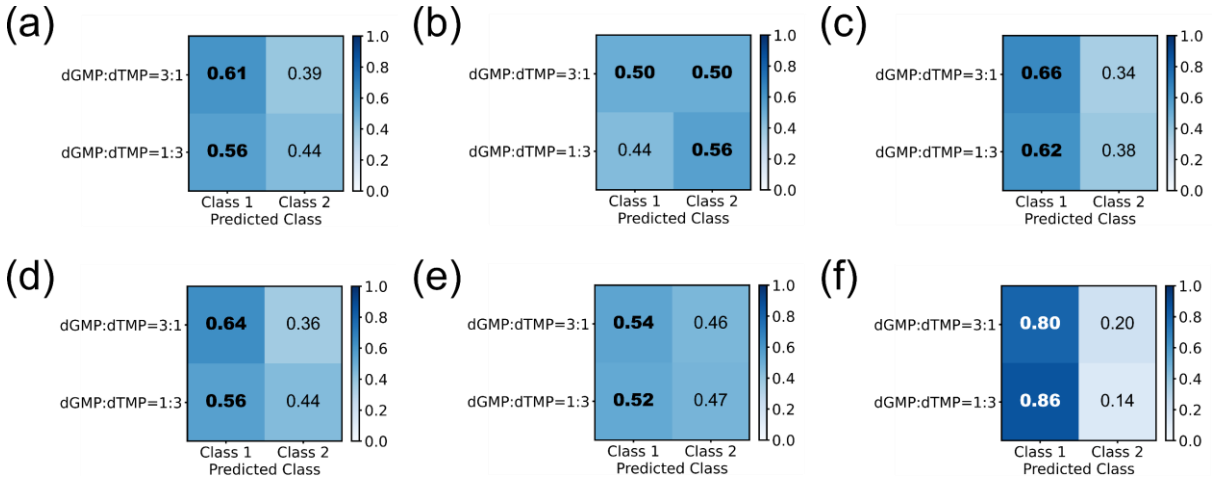

Figure 6.13 Discrimination results with unsupervised machine learning. (a-c) Discrimination results without PCA and (d-f) with PCA. The clustering algorithms are (a d) k-means clustering, (b, e) spectral clustering, and (c, f) GMM clustering.

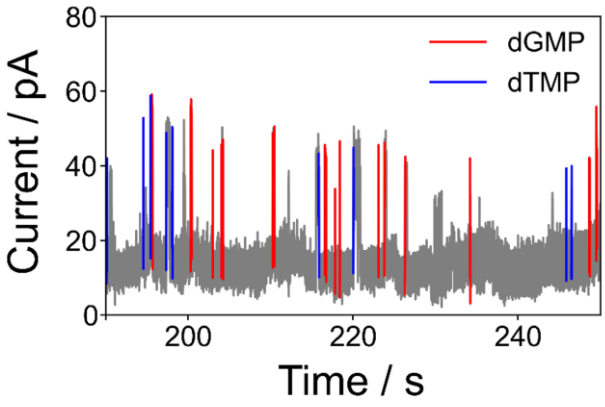

Figure 6.14 The current profile resulting from identifying the signal of each single molecule individually (in dGMP:dTMP=3:1 solution).

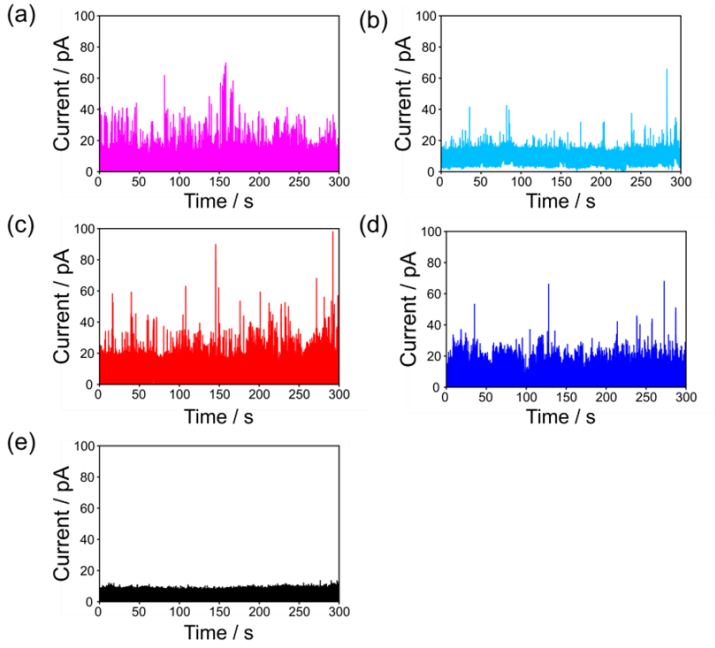

Figure 6.15 The current profiles of each measurement. (a) dGMP solution, (b) dTMP solution, (c) dGMP:dTMP=3:1 solution, (d) dGMP:dTMP=1:3 solution, (e) Milli-Q water as the blank solution. All solutions used in the measurements were prepared using deionized Milli-Q water.

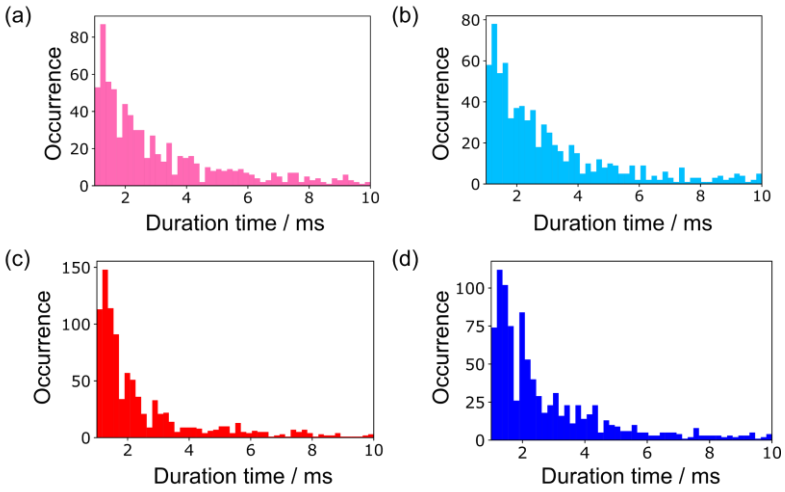

Figure 6.16 Histograms of the duration of the current pulse for each measurement. (a) dGMP solution, (b) dTMP solution, (c) dGMP:dTMP=3:1 solution, (d) dGMP:dTMP=1:3 solution.

6.4 Discussion

The current histograms of dGMP and dTMP in Figure 6.6 show that dGMP levels higher currents than dTMP. However, when focusing on the individual current signals rather than the histogram, dGMP did not always show a higher current than dTMP. The machine learning algorithm used in this study discriminates between the molecules' signals based on the level of current and the shape of the pulse. Figure 6.17 shows I_p histograms representing the discrimination results of signals obtained from solutions with dGMP:dTMP ratios of 3:1 and 1:3. In both histograms, the pink bars represent the histogram of the current predicted by dGMP and the blue bars represent the histogram of the current predicted by dTMP. The two Ip histograms show that relatively high current data is predicted by dGMP and relatively low current data is predicted by dTMP. The histograms show that the UUC method predicts the mixing ratio of the molecules and that dGMP has a higher conductance than dTMP, which is in the same trend as the pure solution measurements. In particular, the new method determines the concentration ratio from the mixed solution measurement of two molecules whose concentration ratio is unknown. This new method can be applied to a variety of molecular detection methods. For example, the technique can be applied to determine the concentration ratio of a foreign substance in a biological sample by comparing it to a normal sample and a positive/negative sample with a control that promotes or inhibits the molecule of interest. In addition, the concentration of the molecule of interest can be determined by comparing a sample with an unknown concentration to a sample to which a reference sample has been added.

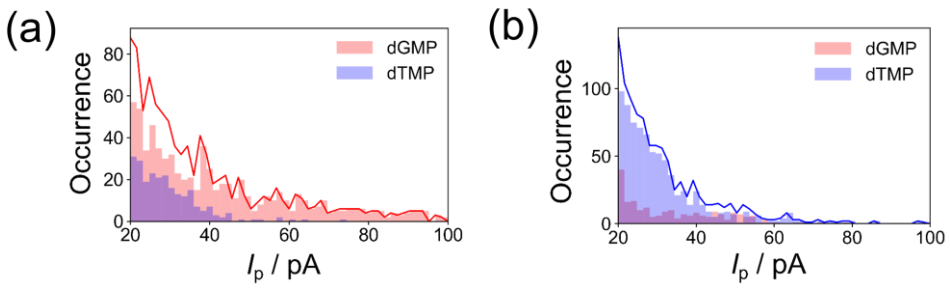

Figure 6.17 I_p histograms based on the identification results of the (a) dGMP:dTMP=3:1 and (b) dGMP:dTMP=1:3 solutions, respectively. The red and blue bars represent the histograms predicted as dGMP and dTMP, respectively. The solid lines represent the sum of the two histograms.

6.6 Conclusion

In this study, I proposed and demonstrated the feasibility of a new method for discriminating molecules using only single-molecule measurement data in mixture solutions, which complemented conventional single-molecule discrimination methods. The new method discriminated well between molecules without their respective labels from measurement data in mixture solutions without any pre-training on pure solutions of the target molecules, which was achieved by applying the UUC algorithm based on KDE. The main method of this goal was to develop a new method and evaluate its performance. For the molecular discrimination of the target molecules dGMP and dTMP in this study, compared with the conventional method, the new method showed improved accuracy in predicting the composition of the mixed solution. The technique for discrimination of target molecules in mixed solutions developed in this study is expected to be utilized in various fields for various molecules without individual sample training.

Chapter 7. Conclusions

In this thesis, I proposed approaches to expand the scope of single-molecule measurement applications, aiming to enhance its potential for diverse fields. The methodology employed involves measuring single molecules freely diffusing in solution, deviating from the conventional approach of chemically bonding molecules to electrodes. Applying this technique to various molecular analytical methods poses a significant challenge and goal for future research.

The primary focus of this study was categorized into two key areas. First, it introduced new measurement approaches to discern molecules that were challenging to differentiate due to structural similarities. Second, it explored analytical methods to streamline conventional discrimination techniques.

In Chapter 4, the approach of modifying the nano-gap electrodes with appropriate molecules to improve the discrimination accuracy of difficult-to-discriminate molecules was discussed. In this study, the target molecules Asp and Leu were discriminated by the change in current properties due to the difference in interaction with MAA molecules modified on the nano-gap electrodes. Compared to the measurement results using bare nano-gap electrodes, the discrimination accuracy of the two target molecules was significantly improved in the case of molecular modification. In this study, I proposed and achieved a method to discriminate molecules that are difficult to discriminate under normal conditions with high accuracy by modifying nano-gap electrodes with appropriate molecules.

In Chapter 5, an approach to improve the accuracy of molecular discrimination through differences in the behavior of molecules due to changes in the chemical environment was discussed. In this study, molecules were discriminated by the difference in behavior of two target molecules due to changes in pH. The target molecules in this study, L-DOPA and dopamine, showed similar current characteristics due to their similar molecular structures

under neutral conditions at pH 7.6, making discrimination difficult. In the acidic condition of pH 3.0, the current characteristics of the two molecules were significantly different, resulting in improved discrimination accuracy. This study demonstrated that molecules that are difficult to discriminate under normal conditions can be discriminated through differences in molecular behavior by changing the chemical environment.

In Chapter 6, a novel approach was discussed that addresses the potential drawbacks of the conventional single-molecule discrimination method and allows for broader application of single-molecule measurements. The conventional method requires training from measured data of samples containing only the target molecule before analyzing practical samples. Since this training process can sometimes limit the analysis of samples, a new method was developed that utilizes the UUC algorithm based on KDE to analyze the molecular composition of samples directly from data of mixed solutions only. It is expected that this method can be developed into a technology that can accurately predict the composition of a solution even in a solution containing a variety of molecules.

This thesis demonstrates the potential for advancing the field of single-molecule discrimination and expanding single-molecule measurement into various fields in the future. For example, applications to the analysis of biomolecules and complex chemical mixtures are expected to revolutionize various fields such as disease diagnosis, environmental monitoring, and drug discovery. To cultivate these valuable and innovative fields, it is crucial to validate the methodology developed in this study under diverse environmental conditions, demonstrating its efficiency and universality. Further research, including diversification of experimental conditions and real-world applications, will be necessary. In addition, the accuracy and efficiency of data analysis can be improved through the improvement and optimization of machine learning algorithms. This is poised to significantly impact various scientific and industrial fields where precise molecular-level analysis is crucial. It is also important to develop the methodology of this study in an affordable and easy-to-use form.

In conclusion, the research presented in this thesis paves the way for new developments in molecular analytical methods, and the results are expected to make a significant contribution to the future development of this field along with the advancement of single-molecule measurement techniques. In particular, the approach and results of this research pave the way

for further research towards practical and real-world applications, and suggest new pathways for scientific inquiry and technological innovation.

Publications

[1] Aeyeon Kang, **Jiho Ryu**, Jisu Lee, Seunghun Kim, Cho Yeon Lee, Wan Soo Yun, (2021), "Calibration curve-free electrochemical quantitation by micro-nano multi-scale gap devices", *Microchimica Acta*, *188*(6), 1-7. (A. Kang and J. Ryu contributed equally to this work)

[2] **Jiho Ryu**, Yuki Komoto, Takahito Ohshiro, Masateru Taniguchi, (2022), "Single-Molecule Classification of Aspartic Acid and Leucine by Molecular Recognition through Hydrogen Bonding and Time-Series Analysis", *Chemistry–An Asian Journal*, *17*(13), e202200179.

[3] **Jiho Ryu**, Yuki Komoto, Takahito Ohshiro, Masateru Taniguchi, (2023), "Direct Biomolecule Discrimination in Mixed Samples using Nanogap-Based Single-Molecule Electrical Measurement", *Scientific Reports*, *13*(1), 9103.

[4] Yuki Komoto, **Jiho Ryu**, Masateru Taniguchi, (2023), "Machine learning and analytical methods for single-molecule conductance measurements", *Chemical Communications*, *59*, 6796-6810.

[5] Yuki Komoto, **Jiho Ryu**, Masateru Taniguchi, (2024), "Total variation denoising-based method of identifying the states of single molecules in break junction data", *Discover Nano*, Accepted.

Conference Presentations

A. Oral presentations

[1] **Single-Molecule Classification Based on Intermolecular Hydrogen Bonding by Modified Nano-Gap**

Jiho Ryu, Yuki Komoto, Takahito Ohshiro, Masateru Taniguchi.

The 69th Spring meeting, Japan Society of Applied Physics (JSAP), Mar., 2022, Sagamihara Campus, Aoyama Gakuin University, Kanagawa, Japan and Online.

[2] **Single amino acid molecule discrimination by molecular modification of nano-gap and machine learning analysis**

Jiho Ryu, Yuki Komoto, Takahito Ohshiro, Masateru Taniguchi.

The 103th Annual Meeting, The Chemical Society of Japan (CSJ), Mar., 2023, Noda Campus, Tokyo University of Science, Chiba, Japan.

[3] **Direct Discrimination of Bio-Molecules in Mixture by Single-Molecule Measurement**

Jiho Ryu, Yuki Komoto, Takahito Ohshiro, Masateru Taniguchi.

The 84th Autumn Meeting, Japan Society of Applied Physics (JSAP), Sep., 2023, Kumamoto-jo Hall and other surrounding facilities, Kumamoto, Japan & Online.

B. Poster presentations

[1] **Single-Molecule Classification Based on Intermolecular Hydrogen Bond by Modified Nano-Gap and Time Series Analysis**

Jiho Ryu, Yuki Komoto, Takahito Ohshiro, Masateru Taniguchi.

The 25th SANKEN International Symposium, Jan., 2022, Online.

[2] **Single-Molecule Classification Based on Intermolecular Hydrogen Bond by Modified Nano-Gap**

Jiho Ryu, Yuki Komoto, Takahito Ohshiro, Masateru Taniguchi.

The 13th Joint Departmental Meeting, Korean Scientists and Engineers Association in Japan (KSEAJ), Mar., 2022, Online.

[3] **Single-molecule Classification of Aspartic Acid and Leucine amino acids using Nano-gap Modification and Machine Learning**

Jiho Ryu, Yuki Komoto, Takahito Ohshiro, Masateru Taniguchi.

The 16th Annual Meeting, Japan Society for Molecular Science, Sep., 2022, Yagami campus, Keio University, Yokohama, Japan.

[4] **Single-molecule Classification of Aspartic Acid and Leucine by Molecular Modification of Nano-gap and Machine Learning**

Jiho Ryu, Yuki Komoto, Takahito Ohshiro, Masateru Taniguchi.

The 83rd Autumn Meeting, The Japan Society of Applied Physics (JSAP), Sep., 2022, Kawauchi-Kita Campus, Tohoku University, Sendai, Japan.

[5] **Single-Molecule L-DOPA/Dopamine Discrimination using Nano-Gap Electrode under Acidic Condition**

Jiho Ryu, Yuki Komoto, Takahito Ohshiro, Masateru Taniguchi.

The 26th SANKEN International Symposium, Jan., 2023, SANKEN, Osaka University, Suita, Japan.

References

1. Aradhya, S. V. & Venkataraman, L. Single-molecule junctions beyond electronic transport. *Nat. Nanotechnol.* **8**, 399–410 (2013).

2. Evers, F., Korytár, R., Tewari, S. & Van Ruitenbeek, J. M. Advances and challenges in single-molecule electron transport. *Rev. Mod. Phys.* **92**, 35001 (2020).

3. Song, H., Reed, M. A. & Lee, T. Single molecule electronic devices. *Adv. Mater.* **23**, 1583–1608 (2011).

4. Bai, J., Li, X., Zhu, Z., Zheng, Y. & Hong, W. Single-Molecule Electrochemical Transistors. *Adv. Mater.* **33**, 1–20 (2021).

5. Huang, C., Rudnev, A. V., Hong, W. & Wandlowski, T. Break junction under electrochemical gating: Testbed for single-molecule electronics. *Chem. Soc. Rev.* **44**, 889–901 (2015).

6. Komoto, Y., Fujii, S., Iwane, M. & Kiguchi, M. Single-molecule junctions for molecular electronics. *J. Mater. Chem. C* **4**, 8842–8858 (2016).

7. Komoto, Y., Fujii, S. & Kiguchi, M. Single-molecule junctions of p molecules. *Mater. Chem. Front.* **2**, 214–218 (2018).

8. Xie, X. *et al.* Single-Molecule Junction: A Reliable Platform for Monitoring Molecular Physical and Chemical Processes. *ACS Nano* **16**, 3476–3505 (2022).

9. Su, T. A. *et al.* Silane and Germane Molecular Electronics. *Acc. Chem. Res.* **50**, 1088–1095 (2017).

10. Xu, B., Xiao, X., Yang, X., Zang, L. & Tao, N. Large gate modulation in the current of a room temperature single molecule transistor. *J. Am. Chem. Soc.* **127**, 2386–2387 (2005).

11. Quek, S. Y. *et al.* Mechanically controlled binary conductance switching of a single-

molecule junction. **4**, (2009).

12. Zhang, J. L. *et al.* Towards single molecule switches. *Chem. Soc. Rev.* **44**, 2998–3022 (2015).

13. Perrin, M. L. *et al.* A gate-tunable single-molecule diode. *Nanoscale* **8**, 8919–8923 (2016).

14. Yamada, R. *et al.* Single-molecule rectifiers based on voltage-dependent deformation of molecular orbitals in carbazole oligomers. *Nanoscale* **10**, 19818–19824 (2018).

15. Fujii, S. *et al.* Rectifying Electron-Transport Properties through Stacks of Aromatic Molecules Inserted into a Self-Assembled Cage. *J. Am. Chem. Soc.* **137**, 5939–5947 (2015).

16. Díez-Pérez, I. *et al.* Rectification and stability of a single molecular diode with controlled orientation. *Nat. Chem.* **1**, 635–641 (2009).

17. Aviram, A. & Ratner, M. A. Molecular rectifiers. *Chem. Phys. Lett.* **29**, 277–283 (1974).

18. Moreland, J. & Ekin, J. W. Electron tunneling experiments using Nb-Sn 'break' junctions. *J. Appl. Phys.* **58**, 3888–3895 (1985).

19. Muller, C. J., van Ruitenbeek, J. M. & de Jongh, L. J. Experimental observation of the transition from weak link to tunnel junction. *Phys. C Supercond. its Appl.* **191**, 485–504 (1992).

20. Letters, Y. L. R. E. V. I. E. & Cd----, A. B. Letters 6. **69**, 140–143 (1992).

21. Bingqian Xu and Nongjian J. Tao. Single-Molecule Resistance Measured by Repeated Formation of Molecular Junctions. *Science (80-.).* **301**, 1221–1223 (2003).

22. Liu, J., Huang, X., Wang, F. & Hong, W. Quantum Interference Effects in Charge Transport through Single-Molecule Junctions: Detection, Manipulation, and Application. *Acc. Chem. Res.* **52**, 151–160 (2019).

23. Cui, L. *et al.* Thermal conductance of single-molecule junctions. *Nature* **572**, 628–633

(2019).

24. Huang, X. *et al.* Electric field–induced selective catalysis of single-molecule reaction. *Sci. Adv.* **5**, 1–8 (2019).

25. Bai, J. *et al.* Anti-resonance features of destructive quantum interference in single-molecule thiophene junctions achieved by electrochemical gating. *Nat. Mater.* **18**, 364–369 (2019).

26. Su, T. A., Neupane, M., Steigerwald, M. L., Venkataraman, L. & Nuckolls, C. Chemical principles of single-molecule electronics. *Nat. Rev. Mater.* **1**, (2016).

27. Hines, T. *et al.* Transition from tunneling to hopping in single molecular junctions by measuring length and temperature dependence. *J. Am. Chem. Soc.* **132**, 11658–11664 (2010).

28. Aragonès, A. C. *et al.* Electrostatic catalysis of a Diels – Alder reaction. *Nature* 8–11 (2016)

29. Yang, C. *et al.* Unveiling the full reaction path of the Suzuki–Miyaura cross-coupling in a single-molecule junction. *Nat. Nanotechnol.* **16**, 1214–1223 (2021).

30. Taniguchi, M. *et al.* Dependence of single-molecule conductance on molecule junction symmetry. *J. Am. Chem. Soc.* **133**, 11426–11429 (2011).

31. Tsutsui, M., Taniguchi, M., Yokota, K. & Kawai, T. Identifying single nucleotides by tunnelling current. *Nat. Nanotechnol.* **5**, (2010).

32. Li, Y. *et al.* Detection and identification of genetic material via single-molecule conductance. *Nat. Nanotechnol.* **13**, 1167–1173 (2018).

33. Ohshiro, T., Tsutsui, M., Yokota, K. & Taniguchi, M. Quantitative analysis of DNA with single-molecule sequencing. *Sci. Rep.* **8**, 8517 (2018).

34. Ohshiro, T. *et al.* Single-molecule electrical random resequencing of DNA and RNA. *Sci. Rep.* **2**, (2012).

35. Ohshiro, T. *et al.* Single-molecule RNA sequencing for simultaneous detection of m6A

and 5mC. *Sci. Rep.* **11**, 1–10 (2021).

36. Ohshiro, T. *et al.* Direct observation of DNA alterations induced by a DNA disruptor. *Sci. Rep.* **12**, 1–9 (2022).

37. Ohshiro, T. *et al.* Detection of post-translational modifications in single peptides using electron tunnelling currents. *Nat. Nanotechnol.* **9**, 835–840 (2014).

38. Hihath, J. & Tao, N. Rapid measurement of single-molecule conductance. *Nanotechnology* **19**, (2008).

39. Ruiz, M. P. *et al.* Bioengineering a Single-Protein Junction. *J. Am. Chem. Soc.* **139**, 15337–15346 (2017).

40. Zhang, B. *et al.* Role of contacts in long-range protein conductance. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 5886–5891 (2019).

41. Zhang, B. *et al.* Observation of giant conductance fluctuations in a protein. *Nano Futur.* **1**, (2017).

42. Komoto, Y. *et al.* Time-resolved neurotransmitter detection in mouse brain tissue using an artificial intelligence-nanogap. *Sci. Rep.* **10**, 1–7 (2020).

43. Nishino, T., Shiigi, H., Kiguchi, M. & Nagaoka, T. Specific single-molecule detection of glucose in a supramolecularly designed tunnel junction. *Chem. Commun.* **53**, 5212–5215 (2017).

44. Zwolak, M. & Di Ventra, M. Colloquium: Physical approaches to DNA sequencing and detection. *Rev. Mod. Phys.* **80**, 141–165 (2008).

45. Di Ventra, M. & Taniguchi, M. Decoding DNA, RNA and peptides with quantum tunnelling. *Nat. Nanotechnol.* **11**, 117–126 (2016).

46. Zwolak, M. & Di Ventra, M. Electronic signature of DNA nucleotides via transverse transport. *Nano Lett.* **5**, 421–424 (2005).

47. Tanaka, H. & Kawai, T. Partial sequencing of a single DNA molecule with a scanning tunnelling microscope. *Nat. Nanotechnol.* **4**, 518–522 (2009).

48. Brown, K. A., Brittman, S., Maccaferri, N., Jariwala, D. & Celano, U. Machine Learning in Nanoscience: Big Data at Small Scales. *Nano Lett.* **20**, 2–10 (2020).

49. Mater, A. C. & Coote, M. L. Deep Learning in Chemistry. *J. Chem. Inf. Model.* (2019)

50. Meuwly, M. Machine Learning for Chemical Reactions. *Chem. Rev.* **121**, 10218–10239 (2021).

51. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

52. Bro-Jørgensen, W., Hamill, J. M., Bro, R. & Solomon, G. C. Trusting our machines: validating machine learning models for single-molecule transport experiments. *Chem. Soc. Rev.* **51**, 6875–6892 (2022).

53. Taniguchi, M. Combination of Single-Molecule Electrical Measurements and Machine Learning for the Identification of Single Biomolecules. *ACS Omega* **5**, 959–964 (2020).

54. Huang, F. *et al.* Automatic classification of single-molecule charge transport data with an unsupervised machine-learning algorithm. *Phys. Chem. Chem. Phys.* **22**, 1674–1681 (2020).

55. Komoto, Y., Ohshiro, T. & Taniguchi, M. Detection of an alcohol-associated cancer marker by single-molecule quantum sequencing. *Chem. Commun.* **56**, 14299–14302 (2020).

56. Taniguchi, M. *et al.* High-Precision Single-Molecule Identification Based on Single-Molecule Information within a Noisy Matrix. *J. Phys. Chem. C* **123**, 15867–15873 (2019).

57. Magyarkuti, A., Balogh, N., Balogh, Z., Venkataraman, L. & Halbritter, A. Unsupervised feature recognition in single-molecule break junction data. *Nanoscale* **12**, 8355–8363 (2020).

58. Liu, B., Murayama, S., Komoto, Y., Tsutsui, M. & Taniguchi, M. Dissecting Time-Evolved Conductance Behavior of Single Molecule Junctions by Nonparametric Machine Learning. *J. Phys. Chem. Lett.* **11**, 6567–6572 (2020).

59. Komoto, Y., Ryu, J. & Taniguchi, M. Machine learning and analytical methods for single-molecule conductance measurements. *Chem. Commun.* 6796–6810 (2023)

60. Landauer, R. Spatial variation of currents and fields due to localized scatterers in metallic conduction. *IBM J. Res. Dev.* **1**, 223–231 (1957).

61. Breit, G. & Wigner, E. Capture of slow electrons. *Phys. Rev.* **49**, 519 (1936).

62. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).

63. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* **13-17-August-2016**, 785–794 (2016).

64. Morikawa, T., Yokota, K., Tsutsui, M. & Taniguchi, M. Fast and low-noise tunnelling current measurements for single-molecule detection in an electrolyte solution using insulator-protected nanoelectrodes. *Nanoscale* **9**, 4076–4081 (2017).

65. Arima, A., Tsutsui, M., Morikawa, T., Yokota, K. & Taniguchi, M. Fabrications of insulator-protected nanometer-sized electrode gaps. *J. Appl. Phys.* **115**, (2014).

66. Gordon, A. H., Martin, A. J. P. & Synge, R. L. M. Partition chromatography in the study of protein constituents. *Biochem. J.* **37**, 79–86 (1943).

67. Zhang, D. *et al.* Raman Detection of Proteomic Analytes. *Anal. Chem.* **75**, 5703–5709 (2003).

68. Tuma, R. Raman spectroscopy of proteins: From peptides to large assemblies. *J. Raman Spectrosc.* **36**, 307–319 (2005).

69. Jenkins, A. L., Larsen, R. A. & Williams, T. B. Characterization of amino acids using Raman spectroscopy. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **61**, 1585–1594 (2005).

70. Li, D. *et al.* Peroxynitrite Activatable NIR-II Fluorescent Molecular Probe for Drug-Induced Hepatotoxicity Monitoring. *Anal. Chem.* **91**, 4771–4779 (2019).

71. Nishino, T., Hayashi, N. & Bui, P. T. Direct Measurement of Electron Transfer through a Hydrogen Bond between Single Molecules. 1–4 (2013)

72. Chaki, N. K. & Vijayamohanan, K. Self-assembled monolayers as a tunable platform for biosensor applications. *Biosens. Bioelectron.* **17**, 1–12 (2002).

73. Bain, C. D. *et al.* Formation of Monolayer Films by the Spontaneous Assembly of Organic Thiols from Solution onto Gold. *J. Am. Chem. Soc.* **111**, 321–335 (1989).

74. Brust, M., Walker, M., Bethell, D., Schiffrin, D. J. & Whyman, R. Synthesis of Thiol-derivatised Gold Nanoparticles in. 801–802 (2000).

75. Daniel, M. C. & Astruc, D. Gold Nanoparticles: Assembly, Supramolecular Chemistry, Quantum-Size-Related Properties, and Applications Toward Biology, Catalysis, and Nanotechnology. *Chem. Rev.* **104**, 293–346 (2004).

76. Xue, Y., Li, X., Li, H. & Zhang, W. Quantifying thiol-gold interactions towards the efficient strength control. *Nat. Commun.* **5**, (2014).

77. Min, B. K., Alemozafar, A. R., Biener, M. M., Biener, J. & Friend, C. M. Reaction of Au(111) with Sulfur and Oxygen: Scanning Tunneling Microscopic Study. *Top. Catal.* **36**, 77–90 (2005).

78. Grandbois, M., Beyer, M., Rief, M., Clausen-Schaumann, H. & Gaub, H. E. How strong is a covalent bond. *Science (80-.).* **283**, 1727–1730 (1999).

79. Sugimoto, Y. *et al.* Quantum degeneracy in atomic point contacts revealed by chemical force and conductance. *Phys. Rev. Lett.* **111**, 1–5 (2013).

80. Geng, H., Yin, S., Chen, K. Q. & Shuai, Z. Effects of intermolecular interaction and molecule-electrode couplings on molecular electronic conductance. *J. Phys. Chem. B* **109**, 12304–12308 (2005).

81. Chen, F., Li, X., Hihath, J., Huang, Z. & Tao, N. Effect of anchoring groups on single-molecule conductance: Comparative study of thiol-, amine-, and carboxylic-acid-terminated molecules. *J. Am. Chem. Soc.* **128**, 15874–15881 (2006).

82. Zotti, L. A. *et al.* Can one define the conductance of amino acids? *Biomolecules* **9**, 1–13 (2019).

83. Scatena, L. F., Brown, M. G. & Richmond, G. L. Water at hydrophobic surfaces:

Weak hydrogen bonding and strong orientation effects. *Science (80-.).* **292**, 908–912 (2001).

84. Sigala, P. A. *et al.* Determination of hydrogen bond structure in water versus aprotic environments to test the relationship between length and stability. *J. Am. Chem. Soc.* **137**, 5730–5740 (2015).

85. Chang, S. *et al.* Tunnelling readout of hydrogen-bonding-based recognition. **4**, 297–301 (2009).

86. Ohshiro, T. *et al.* Electrical Nucleotide Sensor Based on Synthetic Guanine-Receptor-Modified Electrodes. *ChemistrySelect* **3**, 3819–3824 (2018).

87. Postuma, R. B. *et al.* MDS clinical diagnostic criteria for Parkinson's disease. *Mov. Disord.* **30**, 1591–1601 (2015).

88. Ball, N., Teo, W. P., Chandra, S. & Chapman, J. Parkinson's disease and the environment. *Front. Neurol.* **10**, (2019).

89. Bloem, B. R., Okun, M. S. & Klein, C. Parkinson's disease. *Lancet* **397**, 2284–2303 (2021).

90. Agid, Y. Levodopa: Is toxicity a myth? *Neurology* **57**, 858–863 (2001).

91. Tesoro, C. *et al.* An Overview of Methods for L-Dopa Extraction and Analytical Determination in Plant Matrices. *Separations* **9**, 1–26 (2022).

92. Alison Abbott. Levodopa: the story so far. *Nature* **466**, S6–S7 (2010).

93. Zhang, L., Chen, G., Hu, Q. & Fang, Y. Separation and determination of levodopa and carbidopa in composite tablets by capillary zone electrophoresis with amperometric detection. *Anal. Chim. Acta* **431**, 287–292 (2001).

94. Zhao, S., Bai, W., Wang, B. & He, M. Determination of levodopa by capillary electrophoresis with chemiluminescence detection. *Talanta* **73**, 142–146 (2007).

95. Chen, X., Zhang, J., Zhai, H., Chen, X. & Hu, Z. Determination of levodopa by capillary zone electrophoresis using an acidic phosphate buffer and its application in

the analysis of beans. *Food Chem.* **92**, 381–386 (2005).

96. Chou, Y. C., Shih, C. I., Chiang, C. C., Hsu, C. H. & Yeh, Y. C. Reagent-free DOPA-dioxygenase colorimetric biosensor for selective detection of L-DOPA. *Sensors Actuators, B Chem.* **297**, 126717 (2019).

97. Baron, R., Zayats, M. & Willner, I. Dopamine-, L-DOPA-, adrenaline-, and noradrenaline-induced growth of Au nanoparticles: Assays for the detection of neurotransmitters and of tyrosinase activity. *Anal. Chem.* **77**, 1566–1571 (2005).

98. Reddy, N. R., Rhodes, S. & Fang, J. Colorimetric Detection of Dopamine with J Aggregate NanotubeIntegrated Hydrogel Thin Films. *ACS Omega* **5**, 18198–18204 (2020).

99. Vandaveer IV, W. R., Woodward, D. J. & Fritsch, I. Redox cycling measurements of a model compound and dopamine in ultrasmall volumes with a self-contained microcavity device. *Electrochim. Acta* **48**, 3341–3348 (2003).

100. Sandeep, S. *et al.* Electrochemical detection of L-dopa using crude Polyphenol oxidase enzyme immobilized on electrochemically reduced RGO-Ag nanocomposite modified graphite electrode. *Mater. Sci. Eng. B* **232**–**235**, 15–21 (2018).

101. Ahmadi-Kashani, M. & Dehghani, H. A novel selective ternary platform fabricated with MgAl-layered double hydroxide/NiMn2O4 functionalized polyaniline nanocomposite deposited on a glassy carbon electrode for electrochemical sensing of levodopa. *Colloids Surfaces B Biointerfaces* **194**, 111134 (2020).

102. Ji, D. *et al.* Smartphone-based differential pulse amperometry system for real-time monitoring of levodopa with carbon nanotubes and gold nanoparticles modified screen-printing electrodes. *Biosens. Bioelectron.* **129**, 216–223 (2019).

103. Li, J. *et al.* Graphene-Au nanoparticles nanocomposite film for selective electrochemical determination of dopamine. *Anal. Methods* **4**, 1725–1728 (2012).

104. Li, J. *et al.* Synthesis of a manganese dioxide nanorod-anchored graphene oxide composite for highly sensitive electrochemical sensing of dopamine. *Analyst* **145**,

3283–3288 (2020).

105. Goud, K. Y. *et al.* Wearable Electrochemical Microneedle Sensor for Continuous Monitoring of Levodopa: Toward Parkinson Management. *ACS Sensors* **4**, 2196–2204 (2019).

106. Hu, M. & Fritsch, I. Application of Electrochemical Redox Cycling: Toward Differentiation of Dopamine and Norepinephrine. *Anal. Chem.* **88**, 5574–5578 (2016).

107. Yang, Z. *et al.* Carbon nanotubes-functionalized urchin-like In 2S 3 nanostructure for sensitive and selective electrochemical sensing of dopamine. *Microchim. Acta* **177**, 381–387 (2012).

108. Kang, A. *et al.* Calibration curve-free electrochemical quantitation by micro-nano multi-scale gap devices. *Microchim. Acta* **188**, (2021).

109. Kalčec, N. *et al.* Transformation of L-DOPA and Dopamine on the Surface of Gold Nanoparticles: An NMR and Computational Study. *Inorg. Chem.* **61**, 10781–10791 (2022).

110. Zhao, Y. *et al.* Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. *Nat. Nanotechnol.* **9**, 466–473 (2014).

111. Yu, P. *et al.* Single-Molecule Tunneling Sensors for Nitrobenzene Explosives. *Anal. Chem.* (2022)

112. Taniguchi, M. *et al.* High-Precision Single-Molecule Identification Based on Single-Molecule Information within a Noisy Matrix. *J. Phys. Chem. C* (2019)

113. Furuhata, T. *et al.* Highly Conductive Nucleotide Analogue Facilitates Base-Calling in Quantum-Tunneling-Based DNA Sequencing. *ACS Nano* **13**, 5028–5035 (2019).

114. Yoshida, T., Washio, T., Ohshiro, T. & Taniguchi, M. Classification from positive and unlabeled data based on likelihood invariance for measurement. *Intell. Data Anal.* **25**, 57–79 (2021).

115. Bamberger, N. D. *et al.* Beyond Simple Structure-Function Relationships: The Interplay of Geometry, Electronic Structure, and Molecule/Electrode Coupling in

Single-Molecule Junctions. *J. Phys. Chem. C* **126**, 6653–6661 (2022).

116. Bekyarova, E. *et al.* Electronic properties of single-walled carbon nanotube networks. *J. Am. Chem. Soc.* **127**, 5990–5995 (2005).

117. Kaneko, S. *et al.* Identifying the molecular adsorption site of a single molecule junction through combined Raman and conductance studies. *Chem. Sci.* **10**, 6261–6269 (2019).

118. Li, C. *et al.* Charge transport in single Au | alkanedithiol | Au junctions: Coordination geometries and conformational degrees of freedom. *J. Am. Chem. Soc.* **130**, 318–326 (2008).

Acknowledgments

I am deeply grateful to Professor Masateru Taniguchi, my supervisor, for his unwavering support, guidance, and mentorship throughout my doctoral journey. His expertise, constant encouragement, and dedication have profoundly shaped my research and academic development. I am grateful for the opportunity to learn and grow under his guidance.

I extend my heartfelt appreciation to my dissertation committee members, Professors Yasuhiro Nakazawa and Takuya Matsumoto, for their insightful feedback, constructive criticism, and valuable suggestions. Their expertise significantly contributed to refining and enhancing the quality of my dissertation.

Special thanks to Associate Professors Makusu Tsutsui and Takahito Ohshiro, Assistant Professors Hiroyuki Tanaka and Yuki Komoto, all from the Taniguchi group, for their mentorship, support, and encouragement during my academic journey. The invaluable guidance of Yuki Komoto throughout my academic journey, serving as my research mentor, played a pivotal role in shaping my research pursuits and personal growth.

I appreciate the camaraderie, collaboration, and support of all my colleagues at the Taniguchi group, whose insights and discussions enriched my research experience and personal development.

I am deeply indebted to my family and friends for their unwavering care, encouragement, and sacrifices. Furthermore, I would like to express my gratitude to Professor Wan Soo Yun, my master's thesis supervisor at Sungkyunkwan University, for igniting my passion for research and providing the foundation for my doctoral studies.

Finally, I would also like to thank the Japan Science and Technology Agency (JST) and Osaka University for their financial support through 'Support for Pioneering Research Initiated by the Next Generation (SPRING)'.