Supplementary Information

Key Aurophilic Motif for Robust Quantum-Tunneling-Based Characterization of a Nucleoside Analogue Marker

Takafumi Furuhata¹, Yuki Komoto², Takahito Ohshiro ², Masateru Taniguchi^{*2},

Ryosuke Ueki¹, Shinsuke Sando^{*1,3}

¹Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan. ²The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan ³Department of Bioengineering, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan.

Table of Contents

1. Supplementary figures ····································	
2. Supplementary tables ····································	
3. General information ······S1	1
4. Preparation of gap electrodes S1	1
5. Current measurement of nucleosides and nucleotides	1
6. Statistical analysis and base-assignment	2
7. Enzymatic incorporation of EtdU into DNA samples	5
8. Synthesis and purification of oligonucleotides	6
9. Chemical synthesis ······S18	8
10. References	0

1. Supplementary figures



Fig. S1 Signal profiles of dT and its analogs. (a) Typical time traces of the current for the measurements of dT and dT analogs. Expanded traces of EtdU and dT were also provided in Fig. 2c in the main text. (b) (top) Box plots of current intensity, Ave. (bottom) Box plots of the coefficient of variation, CV. Red and blue lines indicate the 75th and 25th percentiles of the values of dT, respectively. Upper and lower limits of box plots are corresponding to the 90th and 10th percentiles, respectively. (c) Two-dimensional histograms of Ave and CV for each of the nucleosides. X-axis represents the base 10 logarithm of the absolute value of Ave (pA). The depth of the red color reflects the relative frequency of each segment normalized by the highest count in each of the histograms.





Fig. S2 Signal profiles of EtdU and AmdU relative to those of the canonical nucleosides. (a) Chemical structures of canonical nucleosides (dT, dC, dA and dG), EtdU, and AmdU. (b) Box plots of (top) Ave and (bottom) CV for each of the nucleosides. Red and blue lines indicate the 75th and 25th percentiles of the values of dT, respectively. Upper and lower limits of the box plots are corresponding to the 90th and 10th percentiles, respectively. (c) Two-dimensional histograms (2D-histograms) of Ave and CV for each of the nucleosides. X-axis represents the base 10 logarithm of the absolute value of Ave (pA). The depth of the red color reflects the relative frequency of each segment normalized by the highest count in each of the histograms.



Fig. S3. Statistical base-assignment processes. (a) Signal features to characterize the nucleosides. For robust calculation of CV values, the signals with more than 10 data points (1 ms) were adopted for further statistical analysis. In addition, signals that are out of the other thresholds (gray box) were excluded from the analysis because of their abnormalities. (b) Schematic flow of base-assignment processes. The signal data were divided into five segments, four of which were adopted as a training data set and the remainder of which was used as a test data set. The probability density functions (PDFs) of the signal features for each of the nucleosides were estimated by kernel density estimation. Based on the PDFs, the probabilities to be assigned to each of the nucleobase was calculated for each signal in the test data set. When the signal coverage = n (n = 1-50), the final probabilities were estimated by naïve Bayes estimation from each of the probabilities calculated for the n signals randomly picked from the test data set. Then, the nucleoside with the largest probability was assigned as the most probable assignment.



Fig. S4 Enzymatic incorporation of EtdU into DNA samples. (a) The gel image showing the result of polyacrylamide gel electrophoresis of the products from primer extension using dNTP mixes containing either of dTTP or EtdUTP (5-ethynyl-2'-deoxyuridine-5'-*O*-triphosphate) as a genetic alphabet "T". The bands of full-length products increased against the extension time in a similar manner between the both cases. The full-length product was not observed when the dNTP mix lacked dTTP and EtdUTP. (b) PCR amplification of TNF α gene using dNTP mixes containing either of dTTP or EtdUTP. A red square indicates the region corresponding to the target PCR products with 702 bp length. The target PCR products were observed in both cases of dTTP (the 2nd lane from the left end) and EtdUTP (the 3rd lane from the left end) in a similar manner. The PCR products were not observed in the control experiments without dTTP and EtdUTP (the 4th lane from the left end) and without the template TNF α gene (the lane on the right end). The line of the left end corresponds to a DNA ladder marker. These results collectively showed that EtdU is efficiently incorporable into DNA as a genetic alphabet "T".



Fig. S5 Signal profiles of oligonucleotide containing EtdU. (a) Typical time traces of 3-mer oligonucleotides containing EtdU (5'-d(XGT)-3', X = EtdU) (upper panel) and lacking EtdU (5'-d(TGT)-3') (lower panel). Red lines indicate the current plateaus detected in the change point analysis using the "changepoint" package of the R language.¹ The details are described in the "6. Statistical analysis and base-assignment" section. (b) Box plots of Ave values calculated from the current plateaus of the oligonucleotide signals. Upper and lower limits of the box plots are corresponding to the 90th and 10th percentiles, respectively. (c) Box plots of CV values calculated from the current plateaus of the oligonucleotide signals. Upper and lower limits of the box plots are corresponding to the 90th and 10th percentiles, respectively. Fig. S5b and S5c showed that the oligonucleotide containing EtdU provides signals with high Ave and/or CV values more frequently than the oligonucleotide lacking EtdU. This indicates that EtdU in the oligonucleotide also exhibits statistically characteristic signals regarding Ave and CV compared with canonical dT and dG.



Fig. S6 Chemical structures and signal features of *p*-disubstituted benzenes. (a) Chemical structures of *p*-disubstituted benzenes. The methoxy group was adopted as a control motif that has a weak gold-binding ability. (b) 2D plots of 75th percentiles (p75) of Ave (*x*-axis) and CV (*y*-axis) for the *p*-disubstituted benzenes. Error bars indicate the standard errors of the p75 values obtained from three independent measurements. (c) (left panel) Box plots of Ave values for the *p*-disubstituted benzenes. (right panel) Box plots of CV values for the *p*-disubstituted benzenes. Red and blue lines indicate the 75th and 25th percentiles of the values of *p*-dimethoxy benzene (*p*-diOMe-Bz), respectively. Upper and lower limits of box plots are corresponding to the 90th and 10th percentiles, respectively. Both Ave and CV values got larger for the benzenes modified with gold-binding motives (cyano, amino, and ethynyl groups) compared with *p*-diOMe-Bz that lacks a gold-binding exocyclic substituent. This observation is consistent with the nucleoside case, indicating that the positive effects of the gold-binding motives on the Ave and CV values of the molecules. In particular, the ethynyl modification led to the most striking increase of Ave value of the benzene. This can be attributed to the unique chemical property of the carbon-carbon triple bond whose interaction with gold involves multiple types of orbital interactions.²



Fig. S7 1D and 2D statistical identification of EtdU in the binary classification of EtdU and one of the canonical nucleosides. (a) Confusion matrices of base-assignment based on 2D-statistical processing of Ave and CV. Signal coverage = 30. (b) Plots of correct assignment rates for each nucleoside against the signal coverage in 1D-statistical processing of Ave. (c) Plots of correct assignment rates for each nucleoside against the signal coverage in 2D-statistical processing of Ave and CV. Correct assignment rates were converged rapidly within fifty-signal coverage by the 2D signal processing compared with 1D signal processing. These observations indicated that the current fluctuation represented by CV is also an important signal feature in addition to the signal intensity for the accurate identification of the dT analog.

2. Supplementary tables

Table S1	The number of measurement	s and signals	for statistical	analysis of Ave	and CV	for each
molecule						

2D-plots (Fig. 2e)	The number of measurements		
dT	9		
dG	5		
AmdU	4		
CNdU	6		
EtdU	12		
Box plots & 2D-histograms			
(Fig. S1 and S2)	The number of signals		
dT	2057		
dC	dC 1729		
dA	1179		
dG	1507		
AmdU	1156		
CNdU	853		
EtdU	6753		
Box plots & 2D-plots	The number of signals		
(Fig. S6)	(The number of measurements)		
<i>p</i> -diCN-Bz	241 (3)		
<i>p</i> -diCC-Bz	2932 (3)		
<i>p</i> -diNH ₂ -Bz	251 (3)		
<i>p</i> -diOMe-Bz	233 (3)		
2D-histograms & box plots	The number of surrent plateous		
(Fig. 4c and S5)	The number of current plateaus		
5'-d(XGT)-3', X = EtdU	735		
5'-d(TGT)-3'	554		

	Detected form	Calcd.	Found
5'-TGT-3'	$[M+H]^+$	876.2	876.1
5'-XGT- $3'$ (X = EtdU)	$[M+H]^{+}$	886.2	886.0

Table S2 The m/z values of the synthesized oligonucleotides found in MALDI-TOF MS.

3. General information

Reagents and solvents were purchased from standard suppliers and used without further purification. Oligonucleotides were synthesized by the automatic DNA synthesizer (M-2-MX, Nihon Techno Service) using commercially available phosphoramidites (glen research). Triisopropylsilyl-5-ethynyl dU phosphoramidite (TIPS-EtdU phosphoramidite) (Glen Research, 10-1555-90) was used to synthetically incorporate EtdU in the oligonucleotide. Purification of oligonucleotide was performed on High Performance Liquid Chromatography Prominence (Shimadzu) using COSMOSIL 5C18-MS-II packed column (Nacalai, 4.6×150 mm, 5 µm). The concentration of oligonucleotide was determined by UV absorbance at 260 nm measured by NanoDrop ND 1000 spectrophotometer (ThermoFisher Scientific). Data analysis was conducted by original codes constructed in python 3.7 and R 3.6.1.

4. Preparation of gap electrodes

The sensor plate with a gold-nano wire was purchased from SCREEN Holdings Co. Ltd.. The gold gap electrodes were prepared by mechanically controlled break junction (MCBJ) method as described in the literature.³ Briefly, the nano-gap devices were fabricated as follows. First, spin-coating polyimide film was applied as an insulating layer on thin-silicon substrate. A nano-gold wire with a width of several tens of nanometers was formed by electron-beam lithography. SiO₂ film was deposited on the gold-wire by chemical vapor deposition. After patterning, dry-etching was performed to form free-standing gold-wire. According to the SEM observation, the etched depth is about 1 μ m. The free-standing gold bridge in SiO₂ plate was subjected to repetitive three-point bending. After several ten times of bending, the gold junction was mechanically broken to generate a pair of gap electrode, which was monitored by the sudden drop of current flow in the electric circuit. The size of gap electrodes was precisely controlled in real time during the current measurement by fine-tuning of lifting bar using a piezoelectric element.

5. Current measurements of nucleosides, oligonucleotides and *p*-disubstituted benzenes

Tunnel current measurements were conducted using 1 μ M deionized aqueous solutions of the molecules, where concentration indicates the strand concentration for the oligonucleotide. In preparation of nucleoside and *p*-disubstituted benzene samples, the compounds were firstly dissolved in DMSO (1 mM) and then the solution was diluted with deionized water by 1000-fold

to adjust final concentration to 1 μ M. For electrical measurements, a 20- μ L sample solution was dropped onto the center of the sensor plate where the gap electrodes were generated. No electrophoretic force was applied to the sample solution and molecules depended on Brownian motion to access gap electrodes. Gap size was set to 0.6 nm and adjusted by the piezoelectric element during all measurements. Bias voltage between a pair of electrodes was 0.1 V. The time of each set of the measurement was typically 1 h. Measurements were carried out more than three sets for each molecule using different gold gap sensors (**Table S1**).

6. Statistical analysis and base-assignment

Picking and analysis of tunnel current signals

A single signal pulse (S) was defined as the current region from the point which exceeded the first threshold ($\eta_1(k)$) to the point which fell below the second threshold ($\eta_2(k)$). To determine those thresholds, the adaptive threshold analysis was adopted from the literature.⁴ In this process, the baseline current ($\mu(k)$) and the standard deviation of baseline current ($\sigma(k)$) that were used to define the η_1 and η_2 values were recursively determined according to the following equations;

$$\mu(k) = (1 - a) \times \mu(k - 1) + a \times i(k)$$
$$\sigma(k) = (1 - a) \times \sigma(k - 1) + a \times \sqrt{(i(k) - \mu(k))^2}$$

where i(k) is the current value at the *k*-th position of the data point and *a* is a regulatory parameter that controls the influence of i(k). In this experiment, 1/30 was adopted as the *a* value. Using the $\mu(k)$ and $\sigma(k)$ values, the thresholds at the *k*-th position of the data point ($\eta_1(k)$ and $\eta_2(k)$) were defined as follows.

$$\eta_1(k) = \mu(k) + 6 \times \sigma(k)$$
$$\eta_2(k) = \mu(k) + 2 \times \sigma(k)$$

Those values were not updated in the signal region (S) and restarted to be updated just after the data points went below $\eta_2(k)$. The averaged intensity from the baseline current (Ave) and the coefficient of variation of the signal (CV) were calculated from the data points included in each S. Those parameters were adopted as the signal features to characterize the nucleosides in this study. The signals with more than 10 data points (corresponding to 1 ms) were selected for statistical analysis of Ave and CV because short signals are not appropriate for calculation of CV. The signals whose features were out of the thresholds shown in **Fig. S3a** were also excluded from the further

statistical analysis and the base-assignment procedures due to their abnormalities or difficulties to be discriminated from noise signals.

Change point analysis and the detection of current plateaus for oligonucleotides

The multistep signals of the oligonucleotide may correspond to more than one nucleotide in the sequence. To extract the current plateaus that might correspond to each of the single nucleotides trapped between the gap, package "changepoint"¹ of R language was adopted (**Fig. S5a**). Ave and CV values were independently calculated for each of the plateaus and assembled to construct 2D histograms (**Fig. 4c**). Specifically, the "cpt.meanvar" function was used with Pruned Exact Liner Time method (PELT)⁵ to detect the change points of the intensity and the variance in the current traces. In this process, the change points were assigned to the points with less than 2% of the type I error rate by specifying "Asymptotic" as a penalty argument with pen.value = 0.02. Then, the current plateaus with more than 10 data points were adopted for the further statistical analysis of Ave and CV.

Base-assignment to the nucleoside signals (Fig. 3 and S7)

Base-assignment to the nucleoside signals were carried out based on the two-dimensional characterization of the signal features (Ave and CV) and one-dimensional characterization of the signal feature (Ave) in **Fig. 3c** and **S7**. Throughout this process, the Ave (pA) and CV values were converted to the base 10 logarithms. The kernel density estimation of probability density functions of the signal features was conducted based on "gaussian_kde" of the scipy.stats module in the "SciPy 1.0" package.⁶ The general descriptions of the kernel density estimation are provided at "Univariate and bivariate kernel density estimation" of this section. The detailed procedures for discriminating the *m* nucleosides (C₁,…, C_k,…, C_m) are described below.

First, the procedures to omit noise signals was conducted. For this purpose, the signals were categorized into two groups, the signals from background measurements and nucleoside samples. 2D or 1D probability density functions of the signal features (PDFs) were independently constructed by kernel density estimation for background and nucleoside groups. Then, the probabilities to be classified as a background noise and a nucleoside signal were calculated for each signal from the nucleoside group based on the PDFs. The signals that were classified as the background noise with no less than 70% probabilities were excluded from the subsequent base-assignment processes.

The signal data sets of each of the nucleosides $(C_1, \dots, C_k, \dots, C_m)$ were independently divided into five segments, four of which were grouped as a training data set, and the rest of which was a test data set. From the training data, the 2D or 1D probability density functions of signal features were independently estimated for each of the nucleosides using kernel density estimation. Subsequently, the test data signals of each of the nucleosides were randomly picked and grouped by every *n* signals (x_1, \dots, x_n) as one signal set where *n* represents the signal coverage. Then, the probability for each set of the signals to be classified as the *k*-th nucleoside, $p(C_k | x_1, \dots, x_n)$, was calculated according to the naive Bayes estimation as described below;

$$p(C_k|x_1, \cdots, x_n) = \frac{1}{Z}p(C_k)\prod_{i=1}^n p(C_k|x_i)$$

where Z is the normalization constant, $p(C_k)$ is the prior probability (0.5 for binary, 0.33 for ternary and 0.2 for quinary), and $p(C_k|x_i)$ is the probability to be classified as C_k for the *i*-th signal (x_i) of the signal set. Finally, the nucleoside with the largest $p(C_k|x_1, \dots, x_n)$ was assigned to the signal set. The random grouping of the test signals into signal sets and base-assignment to all the signal sets were repeated *n* times when the signal coverage = *n*.

A series of the base-assignment processes described above were repeated five times until all possible segmentations of the five data segments into training and test data sets were tested. Finally, for each of the k nucleosides, the percentage of correct and incorrect assignments was calculated for each nucleoside from the total number of correct and incorrect base-assignments and used collectively to provide a confusion matrix.

Univariate and bivariate kernel density estimation

[Univariate kernel density estimation]

Estimation of 1D density function (f(x)) of Ave is performed by the kernel density estimation described by the following equation.

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - x_i)$$

Here, x is a variate (Ave in this study), n is the number of signals, K(u) is a kernel function, h is a bandwidth parameter of the kernel function and $K_h(t) = K(t/h)/h$. In this study, K(u)corresponds to the gaussian function represented as follows.

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$$

The bandwidth parameter h can be estimated by the following equation according to Silverman's rule of thumb.

$$h = \{\frac{4}{(d+2)n}\}^{\frac{1}{d+4}} \times \sigma$$

Here, *d* is the number of the variables (one in the case of the univariate kernel density estimation), and σ is the standard deviation of the observed values of the variate. By substituting d = 1, *h* can be simply represented as follows.

$$h = 1.06\sigma n^{-\frac{1}{5}}$$

[Bivariate (Multivariate) kernel density estimation]

The kernel density estimation can be extended to the multivariate cases. The general description of the probability density function obtained from multivariate kernel density estimation $(\hat{f}(x))$ are the following equations.

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_1 \cdot \dots \cdot h_d} \widehat{K}\left(\frac{x_1 - x_{i1}}{h_1}, \dots, \frac{x_d - x_{id}}{h_d}\right) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{j=1}^{d} \frac{1}{h_j} K\left(\frac{x_j - x_{ij}}{h_j}\right) \right\}$$
$$\widehat{K}(\boldsymbol{u}) = K(u_1) \cdot \dots \cdot K(u_d)$$
$$\boldsymbol{u} = (u_1, \dots, u_d)$$

where h_k is a bandwidth parameter of the *k*-th variate, K(u) is a kernel function, and *d* is the number of the variables (d = 2, in the case of 2D processing of Ave and CV).

7. Enzymatic incorporation of EtdU into DNA samples

Oligonucleotides

For primer extension, 5'-FAM labeled oligonucleotides (primer) were purchased from Eurofin Genomics and a template DNA was purchased from Fasmac. The sequences are as follows.

Primer : 5'-TAATACGACTCACTATAGGGAGA-3'

Template : 5'-CACTAGGACGGCAGAGGAACATTCTCCCTATAGTGAGTCGTATTA-3' (*Bold "A": A position where the first dT or EtdU is incorporated as a genetic alphabet "T") For PCR amplification of a hTNFα coding sequence (NCBI Reference Sequence: #NM_000594.3), forward and reverse primers were purchased from Fasmac. A plasmid vector coding human TNFα cDNA (pBluescript SKII hTNFα) was obtained from RIKEN. The sequences of primers are the followings. Forward primer : 5'-ATGAGCACTGAAAGCATGATCCG-3' Reverse primer : 5'-TCACAGGGCAATGATCCC-3'

Primer extension

The primer (0.4 μ M) and template (0.4 μ M) were mixed in a 10 μ L of reaction mixture containing 2×reaction buffer for Ex Taq polymerase (Takara Bio) and 1 unit of Ex Taq polymerase (Takara Bio). The primer and template were denatured by heating at 95 °C for 2 minutes and then annealed by cooling to 54 °C at the rate of -0.1 °C/sec. A primer extension reaction was started by adding a 10 μ L of 400 μ M dNTP mix containing dTTP (Sigma-Aldrich) or EtdUTP (abcam) as a genetic alphabet "T" and the reaction mixtures were kept at 54 °C. After a defined time, the reaction was stopped by adding a 20 μ L of loading buffer (7 M urea and 20 mM of EDTA in TBE buffer). A 10 μ L of the resulting mixture was analyzed by 15% denaturing polyacrylamide gel electrophoresis. The FAM-modified primers and extended products were detected by a gel imager.

PCR amplification of human TNFα gene

Human TNF α gene (702 bp) was amplified with PCR using pBluescript SKII hTNF α as a template. PCR was performed in 20 µL total reaction volume with primers (1 µM), dNTP mix containing dTTP or EtdUTP (250 µM), 1×reaction buffer for Ex Taq polymerase (Takara Bio) and 1 unit of Ex Taq polymerase (Takara Bio). The reaction mixture was first heated at 95 °C for 2 minutes and followed by 30 cycles of denaturing (95 °C for 30 seconds), annealing (50 °C for 30 seconds) and elongation (72 °C for 60 seconds) steps.

8. Synthesis and purification of oligonucleotides

Synthesis of the oligonucleotides was performed using the automatic DNA synthesizer (M-2-MX, Nihon Techno Service) on a 1.0 µmole scale (500 Å CPG beads as solid support). Each phosphoramidite was dissolved in anhydrous acetonitrile (ACN) to give 0.1 M solution and it was pushed into the CPG column simultaneously with an equal volume of 0.25 M 4,5-dicyanoimidazole in anhydrous ACN as an activator for the coupling reaction. Coupling time of the TIPS-EtdU phosphoramidite was 3 minutes, almost doubled compared to standard coupling time for phosphoramidites of the canonical nucleotides (dT and dG) to ensure sufficient coupling yield. After the coupling of the phosphoramidite, standard capping and oxidation steps were performed. After all coupling steps were completed, the terminal 4,4'-dimethoxytrityl (DMTr) protecting group was cleaved using a standard DMTr off procedure. Then, the CPG beads were

treated with 28% NH₃ aqueous solution at room temperature for 1 hour to cleave off the oligonucleotide. The resulting solutions were heated at 55 °C for 16 hours. After cooling down, the samples were lyophilized. To deprotect the TIPS protecting group from the EtdU precursor, the oligonucleotide was dissolved in DMF (200 μ L) and heated at 50 °C for 30 minutes. Then, 50 μ L of 1 M tetrabutylammonium fluoride (TBAF) in THF was added to the solution. After the mixture was incubated at 45 °C for 15 minutes, the reaction was quenched by adding 250 μ L of 2 M triethyl ammonium acetate (TEAA) buffer. Then, the solution was supplied to a purification step by the reverse phase HPLC. The product was eluted by 100 mM TEAA (pH 7.0)/ACN (ACN: 2% for 5 min, linear gradient of ACN: 2%–25% for 23 minutes). The elution of the oligonucleotide was lyophilized. The resulting dry oligonucleotide was dissolved in deionized water and characterized by MALDI-TOF MS (**Table S2**).

9. Chemical synthesis

NMR spectra were recorded on an ECS400 spectrometer (JEOL). A residual peak of DMSO (2.50 ppm) in DMSO-d₆ was used as an internal standard for ¹H NMR. A peak of DMSO-d₆ (39.5 ppm) was used as an internal standard for ¹³C NMR. ESI mass spectra (ESI-MS) were measured using a micrOTOF II (Bruker).

Synthesis of C5-modified dUs



Scheme 1. Synthesis of C5-cyano dU (CNdU) and C5-amino dU (AmdU). (a) NH₄Cl (11 eq.), KOH (5.0 eq.), triethylamine (10 eq.) in 60% MeCN aqueous solution, 55 °C, 15 h, 90%. (b) Acetic anhydride (1.0 eq.) in pyridine, r.t., 6 h, 56%. (c) NO⁺BF₄⁻ (6.4 eq.) in DMF, r.t., 30 min, 50%. (d) Pd/C (10%) (0.04 eq.) in MeOH, H₂ atmosphere, r.t., overnight, 56%. (e) 8.8 M NH₃, 40%[v/v] pyridine in water, r.t., 3.5 h, 43%.

Synthesis of CNdU⁷ and AmdU^{8,9} was conducted according to the literature as shown in Scheme 1. 5-Ethynyl-2'-deoxyuridin (EtdU) was purchased from FUJIFILM Wako.

5-Cyano-2'-deoxyuridine (CNdU) (2)

NH₄Cl (0.12 g, 2.3 mmol) and KOH (60 mg, 1.1 mmol) was dissolved in 60% MeCN aqueous solution and triethylamine (TEA) (0.20 g, 2.0 mmol) was added to the solution. The mixture was stirred at 40 °C for 10 minutes and compound 1 (60 mg, 0.20 mmol) was added. The reaction mixture was stirred at 55 °C for 15 hours under N₂ atmosphere and the solvent was removed by evaporation. The crude product was purified by silica gel chromatography (AcOEt:MeOH = 150:1). The product was obtained as white solid (46 mg, 0.18 mmol, 90 %). ¹H NMR (400 MHz, DMSO-d₆): δ 11.99 (br, 1H), 8.82 (s, 1H), 6.01 (dd, *J* = 6.2, 6.2 Hz, 1H), 5.27 (d, *J* = 4.6 Hz, 1H), 5.23 (t, *J* = 5.0 Hz, 1H), 4.23 (m, 1H), 3.81 (m, 1H), 3.70–3.54 (m, 2H), 2.22–2.17 (m, 2H); ¹³C

NMR (100 MHz, DMSO-d₆): δ 160.2, 149.6, 149.1, 114.5, 87.9, 87.8, 85.9, 69.1, 60.3, 40.4; HRMS (ESI) m/z: [M–H]⁻ Calcd for C₁₀H₁₁N₃O₅⁻ 252.0615; found 252.0612.

5-Amino-2'-deoxyuridine (AmdU) (7)

Compound 6 (0.13 g, 0.40 mmol) was dissolved in pyridine (5 mL) and 15 M NH₃ (7 mL) aqueous solution was added to the mixture (final concentration, 8.8 M). After the reaction mixture was stirred at room temperature for 3.5 hours, another portion of 15 M NH₃ (2 mL) was added. The solution was further stirred for 1.5 hours and the solvent was evaporated. The crude product was purified by silica gel chromatography (CH₂Cl₂:MeOH = 92:8). The product was obtained as white solid (56 mg, 0.17 mmol, 43 %). ¹H NMR (400 MHz, DMSO-d₆): δ 11.28 (br, 1H), 6.93 (s, 1H), 6.20 (dd, *J* = 7.1, 7.1 Hz, 1H), 5.22 (d, *J* = 4.1 Hz, 1H), 4.89 (m, 1H), 4.20 (m, 1H), 4.11 (br, 2H), 3.72 (m, 1H), 3.55–3.50 (m, 2H), 2.04–1.98 (m, 2H); ¹³C NMR (100 MHz, DMSO-d₆): δ 160.5, 148.8, 123.1, 114.5, 86.9, 83.4, 70.6, 61.8, 38.7; HRMS (ESI) m/z: [M–H][–] Calcd for C₉H₁₃N₃O₅[–] 242.0771; found 242.0781.

10. References

- Killick, R. & Eckley, I. A. changepoint: an R package for changepoint analysis. *J. Stat.* Softw. 58, 1–19 (2014).
- Nakazumi, T., Kaneko, S., Matsushita, R. & Kiguchi, M. Electric conductance of single ethylene and acetylene molecules bridging between Pt electrodes. *J. Phys. Chem. C* 116, 18250–18255 (2012).
- Tsutsui, M., Taniguchi, M. & Kawai, T. Fabrication of 0.5 nm electrode gaps using selfbreaking technique. *Appl. Phys. Lett.* 93, 163115 (2008).
- 4. Harashima, T. *et al.* Highly reproducible formation of a polymer single-molecule junction for a well-defined current signal. *Angew. Chem. Int. Ed.* **58**, 9109–9113 (2019).
- 5. Killick, R., Fearnhead, P. & Eckley, A. Optimal detection of changepoints with a linear computational cost. *J. Am. Stat. Assoc.* **107**, 1590–1598 (2012).
- Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272 (2020).
- Gillies, A. T., Gai, X. S., Buckwalter, B. L., Fenlon, E. E. & Brewer, S. H. 15N NMR studies of a nitrile-modified nucleoside. *J. Phys. Chem. B* 114, 17136–17141 (2010).
- Kovaliov, M., Weitman, M., Major, D. T. & Fischer, B. Phenyl-imidazolo-cytidine analogues: Structure-photophysical activity relationship and ability to detect single dna mismatch. J. Org. Chem. 79, 7051–7062 (2014).
- Kanamori, T. *et al.* Fluorescence enhancement of oligodeoxynucleotides modified with green fluorescent protein chromophore mimics upon triplex formation. *Org. Biomol. Chem.* 15, 1190–1197 (2017).