Supporting Information

for

A Highly Conductive Nucleotide Analogue Facilitates Base-Calling in Quantum Tunnelling-Based DNA Sequencing

Takafumi Furuhata,[1] Takahito Ohshiro,[2] Gaku Akimoto,[1] Ryosuke Ueki,[1]
Masateru Taniguchi[2*] and Shinsuke Sando[1,3*]

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

Table of Contents

1. Supporting figures

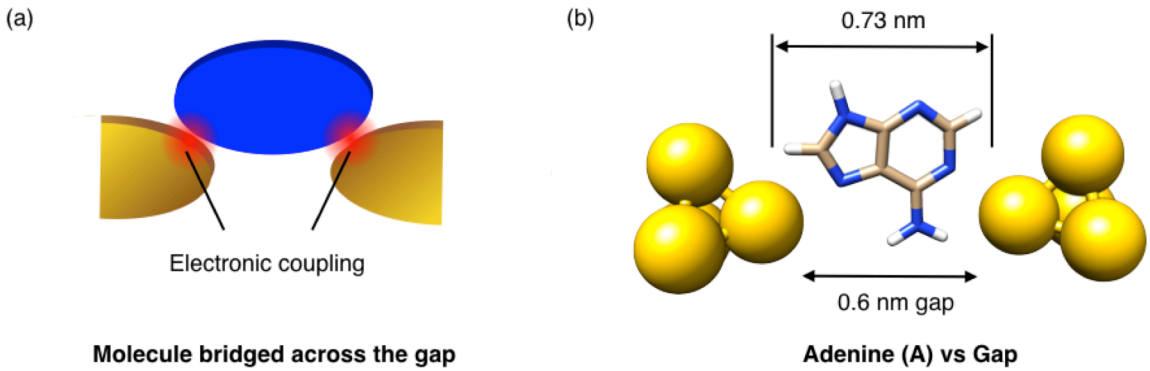


Figure S1. Schematic illustration of electronic coupling between a nucleobase and gap electrodes. (a) Schematic illustration of a molecule bridging across a pair of gap electrodes. This method can detect a molecule that electronically coupled with a pair of gap electrodes even if the size of the molecule is larger than that of the gap. (b) Schematic illustration of adenine (A) passing through the gap. The size of gap (0.6 nm) is smaller than those of canonical purine nucleobases but it was previously shown that aromatic compound could interact with the gap electrodes smaller than its size by electronic coupling between aromatic ring and the metal[1].

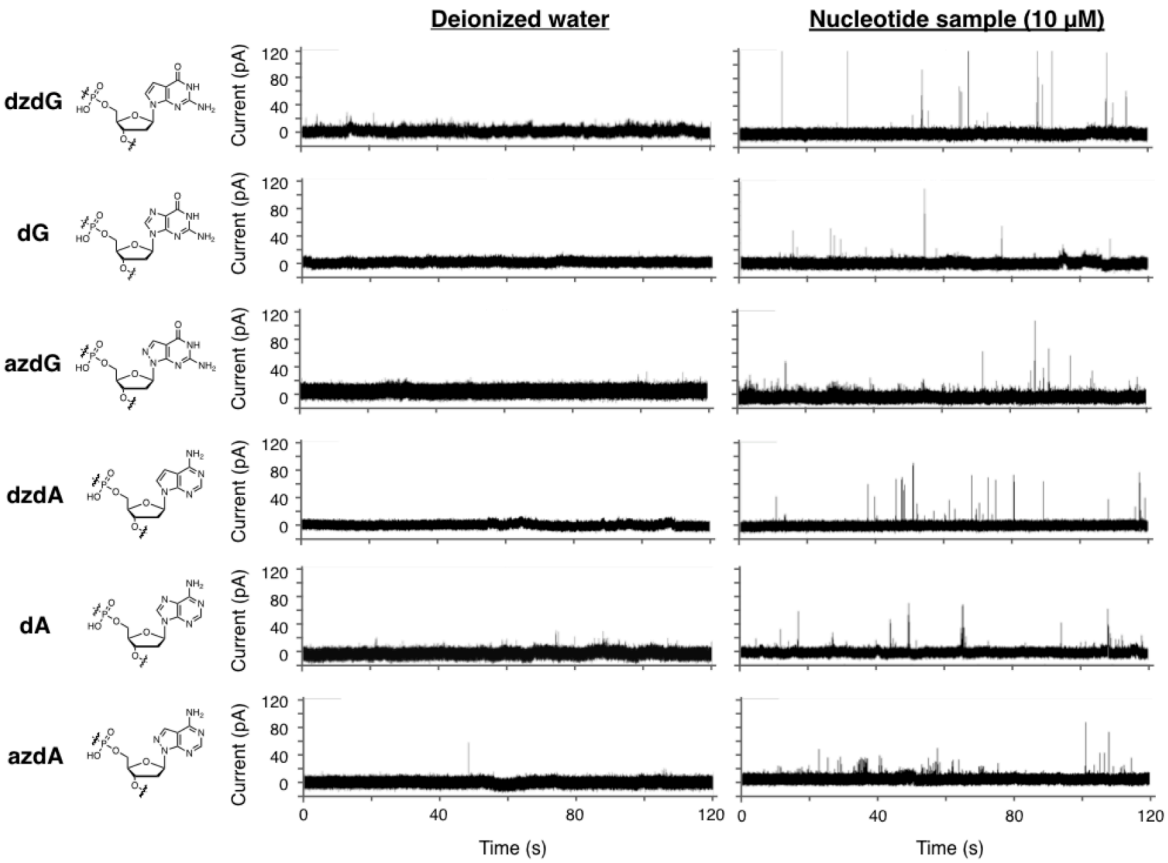


Figure S2. Current measurement of background (deionized water) and sample solutions. Typical current traces of deionized water as background (left panel) and 10 μM of nucleotide solutions (right panel). Current peaks frequently emerged when a nucleotide sample was applied on the sensor plated.

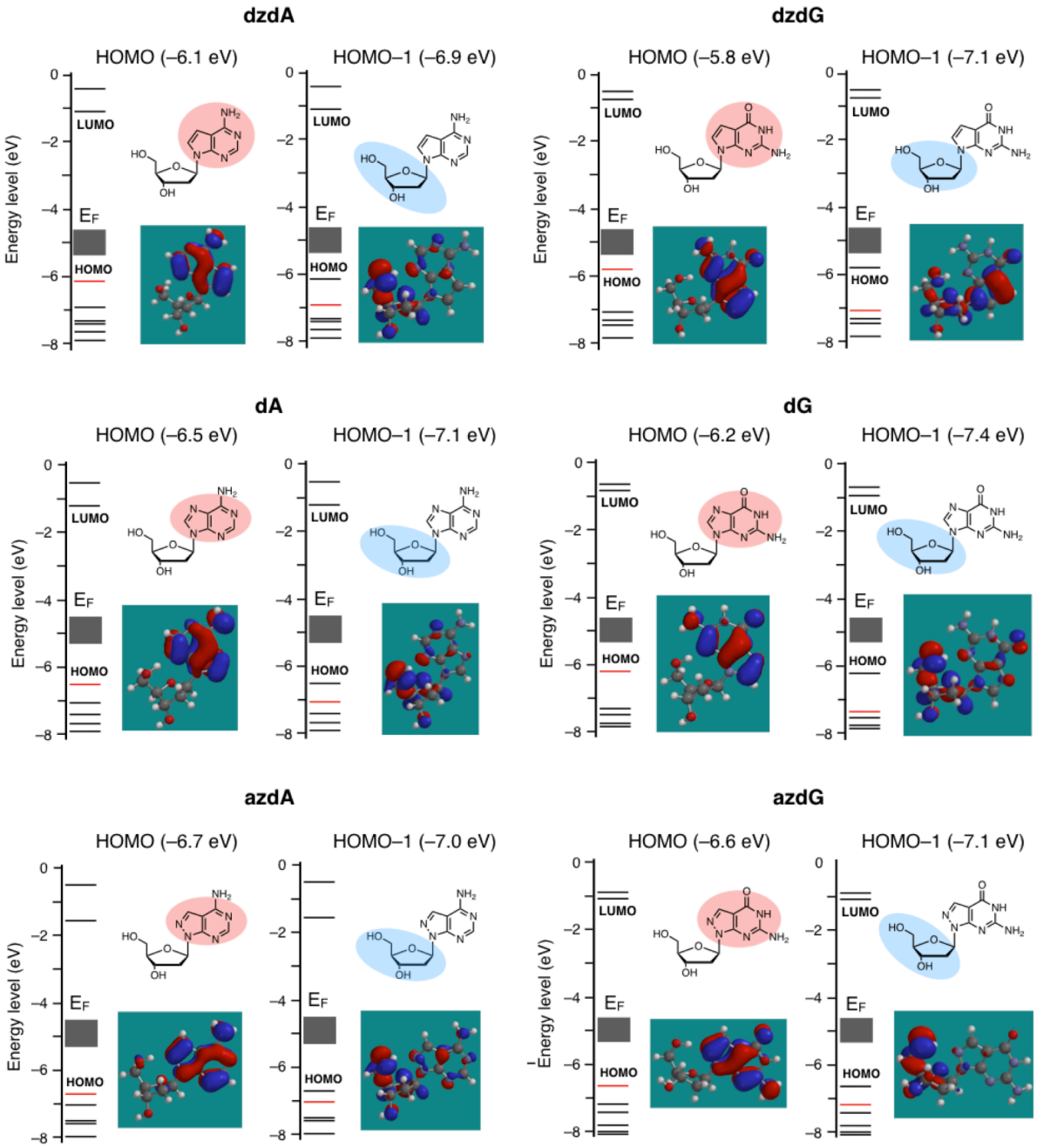


Figure S3. Electron clouds on nucleobases and ribose backbone of nucleosides calculated by density function theory (DFT). Quantum chemical calculation was conducted by B3LYP/6-311+G(d,p)//B3LYP/6-31+G(d). In all cases, electrons at HOMO were localized on the nucleobase. On the other hand, electrons at HOMO–1 (around –7 eV) existed on the ribose backbone. These local electronic states on nucleobase and ribose backbone were supposed to result in multiple conductance distributions of the nucleotide as shown in Figures 3e and 4c.

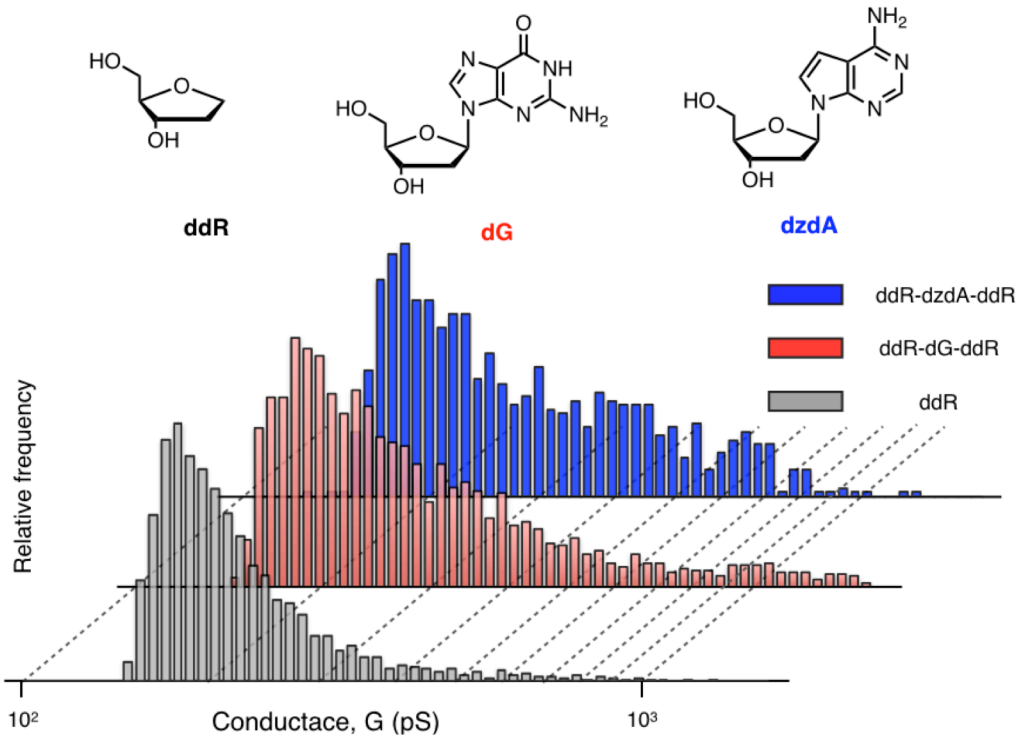
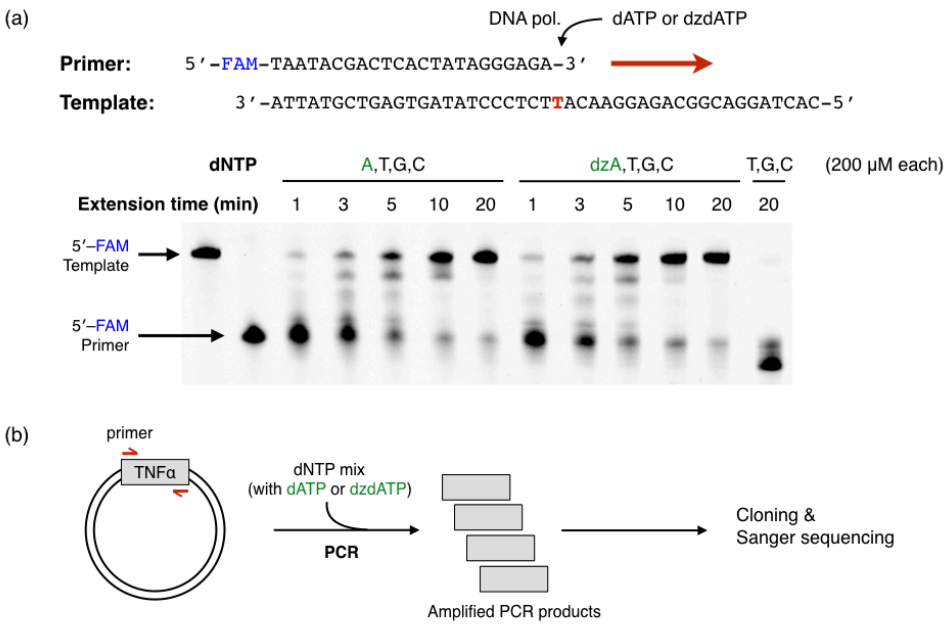
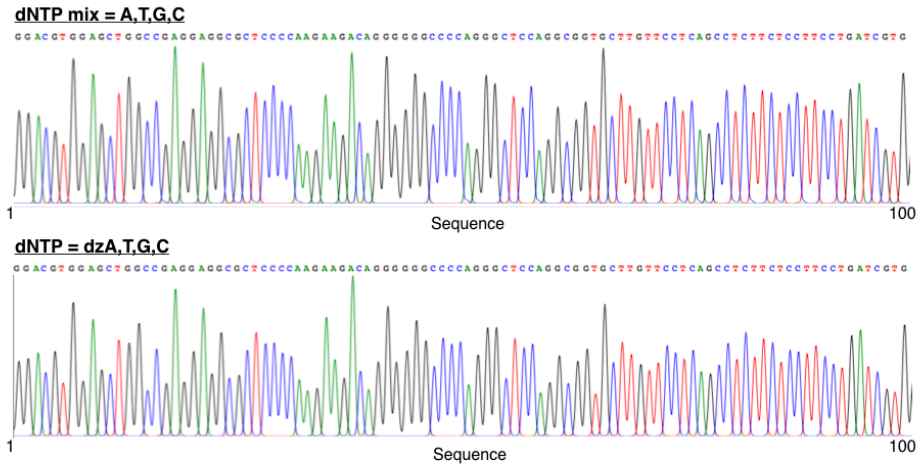


Figure S4. The histograms of molecular conductance of abasic ribose (ddR), oligonucleotides containing dG and dzdA. Upper panel shows the chemical structure of abasic ribose mimic (ddR), dG and dzdA in their nucleoside forms. Lower panel shows conductance histograms taken in a logarithmic scale. The gap size was 0.6 nm and the bias voltage between gap electrodes was 100 mV. Concentration of ddR and oligonucleotides (strand concentration) was 10 μM. The conductance histogram of ddR was limited in lower conductance region compared to those of oligonucleotides containing dG and dzdA. This fact supported the hypothesis that the conductance distribution of oligonucleotides at low conductance region (Figure 3e) resulted from electron transport through ribose backbone (Figures 3g and S3).



Template sequence of amplified region (1st–100th):

5′-GGACGTGGAGCTGGCCGAGGAGGCGCTCCCCAAGAAGACAGGGGGGCCCCAGGGCTCCAGGCGGTGCTTGTTCCTCAGCCTCTTCTCCTTCCTGATCGTG...-3′



[Table] The number of correct base-calling

	A (dzA)	C	G	T	Total
dATP	630 / 630	1120 / 1120	985 / 985	630 / 630	3305 / 3305
dzdATP	630 / 630	1120 / 1120	985 / 985	630 / 630	3305 / 3305

Figure S5. Enzymatic incorporation of dzdA into DNA samples in place of dA. (a) Primer extension to examine the efficiency of enzymatic incorporation of dA or dzdA triphosphate into a DNA template. (Upper panel) Sequences of the template and the primer were designed for a DNA polymerase (Ex Taq) to initiate primer extension reaction by incorporating dA or dzdA that are complementary to the first unpaired thymidine (dT) in the template. (Lower panel) The result of 15% denaturing PAGE analyzing the products generated by 1, 3, 5, 10 and 20 minutes of extension in 20 μL total reaction volume with 200 μM dNTP and Ex Taq polymerase (1 unit). Successful incorporation of the first dA or dzdA resulted in the band emerging at the same position with the 5'-FAM labeled template. (b) (Upper panel) Schematic illustration of the experimental flow to evaluate incorporation of dzdA into a biologically relevant sample. The hTNFα gene (702 bp) was amplified as a template by KOD –Plus- Neo DNA polymerase. PCR product was subsequently supplied to TA cloning and the sequences of inserted gene from 5 colonies were verified with Sanger sequencing. (Middle panel) Partial sequencing chromatograms of PCR products that were amplified using dATP or dzdATP as substrates for a genetic alphabet "A". These chromatograms are corresponding to the first 100 bases of the amplified region that is flanked by forward and reverse primer-binding sites. The sequencing quality was comparable to each other. (Lower panel) A table summarizing the number of correct base-calling in Sanger sequencing of PCR amplified region flanked by two primer-binding sites. There was no detectable mutation in the DNA samples tested in this experiment.

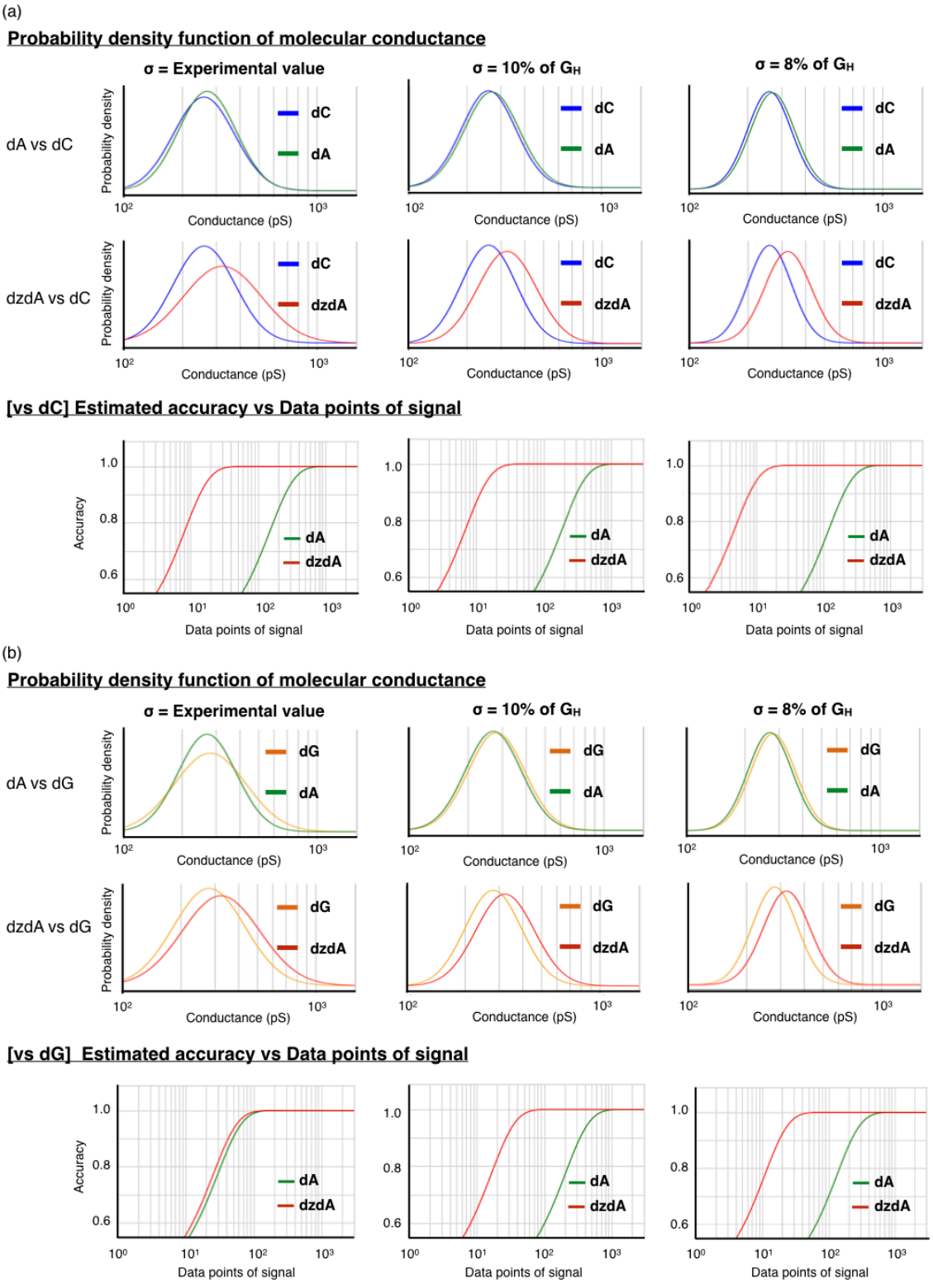


Figure S6. Simulated plots of the accuracy of discrimination of a pair of nucleotides against the required data points of signal. The upper panels show the probability density functions extracted from the Gaussian fitting curves in conductance histograms of each nucleobase (shown in Figure 4c of the main text). The lower graphs show the plot of estimated accuracy of base assignment against the amount of signal points in discrimination of dA (green) or dzdA (red) from (a) dC and (b) dG. This simulation study was conducted using three patterns of probability density function with different standard deviations (σ = experimental values, 10% and 8% of G_H, respectively). The experimental standard deviations (σ) were 10~13% of G_H. The latter two cases were assuming the technological advancement to control the DNA orientation. These sharper σ values modestly improved the accuracy in the case of dzdA but was not effective enough to facilitate the discrimination of dA from dC and dG due to their similarity of the molecular conductance.

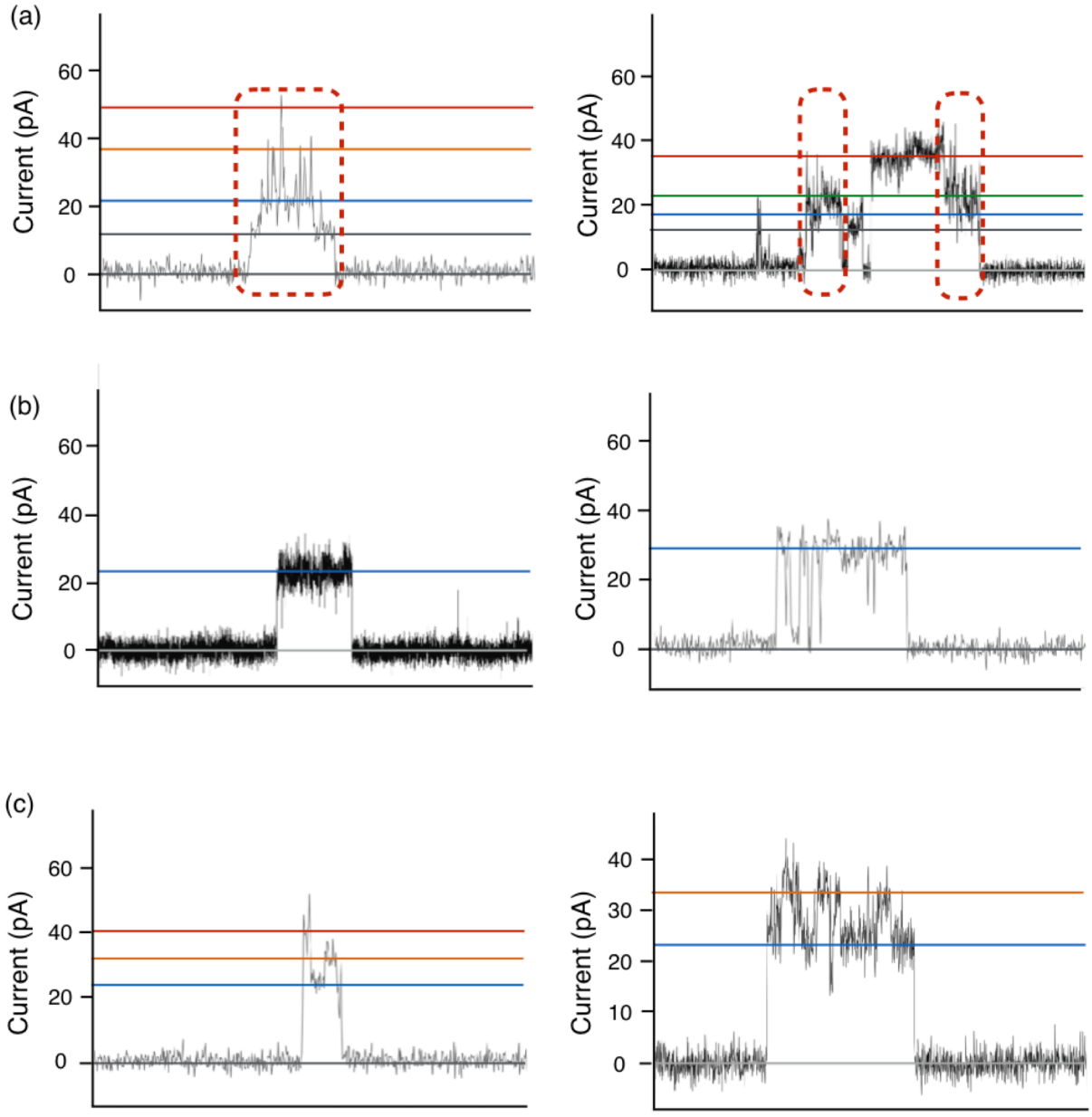


Figure S7. Examples of unideal signals for dzdA-seq. (a) Signals with large current fluctuation. Though multiple levels of current were observed in those signals, current plateaus are difficult to determine due to the fast and large fluctuation of signals. These situations might be caused by too fast passing and large conformational fluctuation of a nucleotide around gap electrodes. This impeded the accurate base assignment to the current trace. Red squares are the regions where the current were especially fluctuated. (b) Mono-level signals. Those signals might indicate that a single nucleobase in an oligonucleotide interacted with gap electrodes and dropped off without passing through the gap electrodes. **c**. Signals with two or three levels of current. Though these signals might be corresponding to the translocation of a part of the oligonucleotide, the length of signals were not sufficient to cover the whole sequence of oligonucleotides. This was probably because the oligonucleotide dropped off in the middle of translocating through the gap electrodes. In all the figures, black and colored lines show apparent current levels of the signals though genetic alphabets could not be assigned due to the large fluctuation of current or insufficient numbers of current levels.

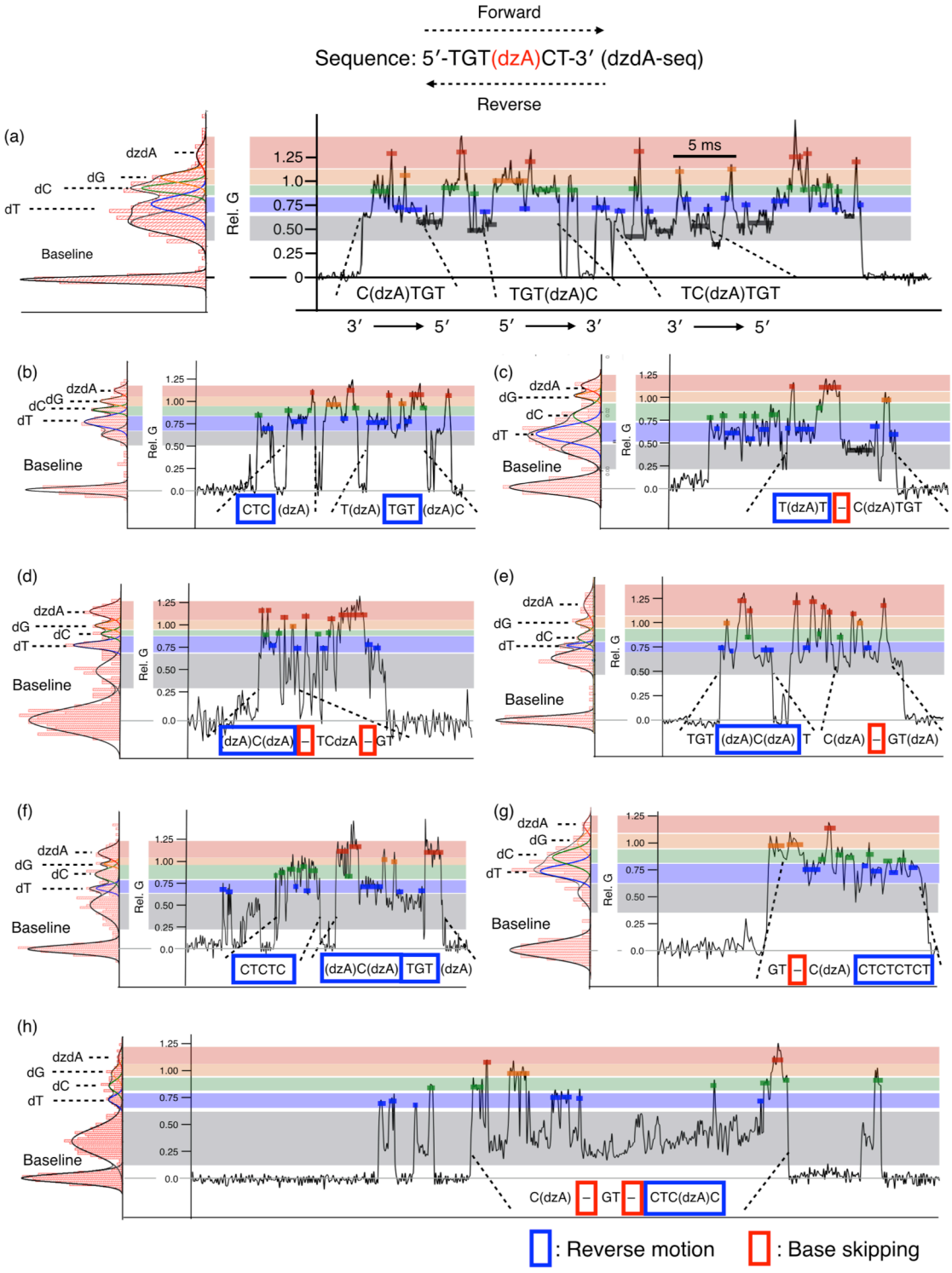


Figure S8. Putative base-assignment in stepwise signals for dzdA-seq. (a)–(h). Stepwise signals for dzdA-seq with signal plateaus that could be assigned to four genetic alphabets. The relative peak conductance values (Rel. G) of the upper four Gaussian fitting curves were the following values; (a) 0.74, 0.90, 1.00 and 1.22, (b) 0.77, 0.90, 1.00 and 1.12, (c) 0.58, 0.78, 1.00 and 1.11, (d) 0.79, 0.91, 1.00 and 1.15, (e) 0.77, 0.85, 1.00 and 1.18 (f) 0.72, 0.88, 1.00 and 1.13, (g) 0.74, 0.86, 1.00 and 1.16 and (h) 0.74, 0.87, 1.00 and 1.15. Putative base assignment to each plateau was conducted based on the probability density functions (equivalent to Gaussian fittings to the current histograms). The alphabets below the current traces are the sequences putatively determined in the base-assignment. These base-assignment processes yielded full or partial sequences of dzdA-seq including base-skipping (red square) and reverse sequence (blue square). It should be noted that the amount of data points assigned to each nucleobase is not sufficient to statistically interpret the signal in the current system. For accurate base-calling, the nucleobases have to be held on the gap electrodes for a time that is sufficient to produce a statistically valid amount of signal data points and each current plateau must be unequivocally associated with each trapping event of a nucleobase.

2. Supporting Tables

Table S1. The number of signals or data points obtained in current measurement of nucleotides for statistical analysis.

Nucleotides (Figure 3)	Number of signals
ddR-azdA-ddR	2238
ddR- dA -ddR	1654
ddR-dzdA-ddR	922
ddR-azdG-ddR	2357
ddR- dG -ddR	2272
ddR-dzdG-ddR	3140

Nucleotides (Figure 4)	Number of signals
ddR- dT -ddR	1725
ddR- dC -ddR	1390
ddR- dA -ddR	1654
ddR- dG -ddR	2272
ddR-dzdA-ddR	922

Hetero-oligonucleotides (Figure 5)	Number of data points
d(TGTACT)	1893
d(TGTdzACT)	10122

*ddR represents the abasic mimic (1,2-dideoxy-D-ribose).

Table S2. The energy levels of frontier orbitals. Energy levels of the frontier orbitals calculated based on density function theory (DFT). Single point calculation was conducted using B3LYP/6-311+G(d,p) after optimization of equilibrium geometry of nucleosides by molecular mechanics (MM) calculation and sequential DFT calculation using B3LYP/6-31+G(d).

	HOMO (eV)	**LUMO (eV)**
dT	–6.92	–1.62
dC	–6.58	–1.53
dA	–6.51	–1.21
dG	–6.24	–0.94
7-deaza-8-aza dA (azdA)	–6.73	–1.59
7-deaza dA (dzdA)	–6.12	–1.07
7-deaza-8-aza dG (azdG)	–6.63	–1.13
7-deaza dG (dzdG)	–5.83	–0.77

Table S3. The m/z values of synthesized oligonucleotides found in MALDI-TOF MS (reflector positive mode).

	detected form	**Calcd.**	**Found**
ddR*-azdA-ddR	$[M+H]^+$	612.15	612.26
ddR-dzdA-ddR	$[M+H]^+$	611.15	611.05
ddR-azdG-ddR	$[M+H]^+$	628.14	628.31
ddR-dzdG-ddR	$[M+H]^+$	627.15	627.05
ddR- dT -ddR	$[M+H]^+$	603.14	602.91
ddR- dC -ddR	$[M+H]^+$	588.14	587.98
ddR- dA -ddR	$[M+H]^+$	612.15	611.87
ddR- dG -ddR	$[M+H]^+$	628.14	628.32

*ddR represents the abasic mimic (1,2-dideoxy-D-ribose).

3. General information

Reagents and solvents for chemical synthesis were purchased from standard suppliers and used without further purification. Hetero-oligonucleotides containing only canonical nucleotides were purchased from Fasmac. A hetero-oligonucleotide containing 7-deaza dA (dzdA) was purchased from GeneDesign. Oligonucleotide containing abasic mimic (1,2-dideoxy-D-ribose, ddR) was synthesized by automatic DNA synthesizer (M-2-MX, Nihon Techno Service) using commercially available phosphoramidites (Glen Research) for canonical nucleotides, 8-aza-7-deaza dA (azdA) and 8-aza-7-deaza dG (azdG) and synthesized phosphoramidites for ddR, dzdA and 7-deaza dG (dzdG). Purification of oligonucleotides 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 oligonucleotides was determined by UV absorbance at 260 nm measured by NanoDrop ND 1000 spectrophotometer (ThermoFisher Scientific). Quantum chemical calculations were conducted using Spartan 16' ParallelSuite (Wavefunction).

4. MO calculations

The structure of a nucleoside was firstly optimized by calculating conformer distributions using Molecular Mechanics (MM) calculation (MMFF). The initial structures for geometry optimization by MM were generated by random bond rotation until one thousand of structures were examined. The most stable structure was employed as an initial structure for further molecular orbital (MO) calculation. Optimization of an equilibrium geometry was conducted based on Density Function Theory (DFT) using B3LYP/6-31+G(d) followed by single point DFT calculation using B3LYP/6-311+G(d,p) to calculate the energy level of molecular orbitals.

5. Preparation of gap electrodes

The gold gap electrodes were prepared by mechanically controlled break junction (MCBJ) method as described in literature[1,2]. 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. SiO_2 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 μm. The free-standing gold bridge in SiO_2 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 electrodes, 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.

6. Measurement of current traces of nucleotides

Tunnel current measurements of nucleotides were conducted using 10 μM of deionized aqueous solutions. Sample solutions were simply dropped (20 μL for each measurement) at the center of the sensor plate where gap electrodes were generated. No electrophoretic force was applied to the sample solution and nucleotides access the gap electrodes depending on Brownian motion. Gap size was set to 0.6 nm and finely tuned by a piezoelectric element during all measurements. Bias voltage between a pair of electrodes was 100 mV. The time for each set of measurement was 10 minutes and it was sequentially followed by another set of measurement. Before total measurement time reached 1.5 hours, the sample solution was replaced with new one. Measurement was repeated until around one thousand signals were observed for each nucleotide.

7. Statistical analysis and base assignment

Signal picking for nucleotides (Figures 3e, 4c, 5b and S4)

For evaluating molecular conductance of nucleotides, the maximum current value in the signal was picked to determine signal intensity (I_p = maximum current value – basal current). In this process, two thresholds were defined to determine the beginning and the end of the signal. The first threshold for the beginning of the signal (Th_1) was (basal current + $Z \times \sigma$) and the second threshold for the end of the signal (Th_2) was (basal current + σ) where the basal current was the simple moving average and σ is the moving standard deviation of the nearest 5000 data points. Z value which was used to determine Th_1 was calculated in each set of measurement (10 minutes) separately as followings: Z = 15 (pA) / STV_{Med} (pA), where STV_{Med} was the median of standard deviation of every 10000 data points (1 second) in the raw current trace. STV_{Med} was typically between 2–3 pA so that Z values were usually within 5–7. The current measurements whose STV_{Med} exceeded 3 pA were excluded from statistical analysis because Z values (or Th_1) get too low to omit noise peaks in such cases. Molecular conductance (G) was calculated by I_p/V where V is the bias voltage (100 mV).

For statistical analysis of current intensity of hetero-oligonucleotide samples, signals that exceeded the threshold (baseline current + $6 \times \sigma$) were picked up. In this case, all data points that exceeded the threshold was collected because signals derived from nucleobases with lower molecular conductance might be ignored in some stepwise signals if only the maximum current was adopted for signal intensity of each signal.

Two Gaussian fitting to the conductance histogram of nucleotides (Figures 3e and 4c)

Each measurement was repeated until around one thousand of signals were obtained for each case. After current intensity was converted to molecular conductance, all the conductance values were combined to make the conductance histogram for each nucleotide. Bin size of $Log_{10}(G)$ was 0.02 where G is the absolute value of conductance (pS). To evaluate molecular conductance of each nucleotide, two Gaussian fittings were applied to each of the conductance histograms which were taken in the logarithmic scale. The integrated fitting curve of the conductance histogram ($F(Log_{10}G)$) is described by the following equation,

$$F(Log_{10}G) = \frac{N_1}{\sqrt{2\pi{\sigma_1}^2}}\exp\left\{-\frac{(Log_{10}G-\mu_1)^2}{2{\sigma_1}^2}\right\} + \frac{N_2}{\sqrt{2\pi{\sigma_2}^2}}\exp\left\{-\frac{(Log_{10}G-\mu_2)^2}{2{\sigma_2}^2}\right\}$$

where $\mu_{1(2)}$ are the peak values of $Log_{10}G$, $\sigma_{1(2)}$ are the standard deviations of $Log_{10}G$ and $N_{1(2)}$ are the areas of two Gaussian distributions. In all the two Gaussian fittings, $(N_1 + N_2) = 0.02 \times S$ where "0.02" means the bin size of the histogram and S is the total number of signals that construct the histogram (Table S1). This fitting condition means that the total area of two Gaussian fitting curves equals that of the histogram of interest. The resulting two peak values ($\mu_{1(2)}$) were converted to conductance values, and they were defined as G_L for the lower value and G_H for the higher value.

Putative base-assignment for stepwise signals in dzdA-seq (Figure S8)

Sequence of oligonucleotides can be stochastically interpreted out of each trace of signals based on the density probability functions that are extracted from each signal trace as described below. (1) The histogram of signal intensity was constructed using all the data points in the current trace extracted from raw data. We hypothesized that the current histogram was a mixture of current distributions given by baseline, nucleobases (T, C, G, and dzA) and low conductive backbone of DNA. Based on this hypothesis, the current histogram can be considered as the sum of the probability density functions of current output of those factors. (2) Multiple Gaussian fittings were applied to the histogram to extract the probability density function corresponding to baseline, nucleobases and low conductive DNA backbone. Since it was indicated that the distribution of "nucleobase" conductance follows a single fitting curves (Figures 3e and 4c), the probability density functions corresponding to four nucleobases conductance can be extracted by the multiple Gaussian fittings. In this procedure, the boundary condition was set for the upper four fitting curves that were putatively assigned to four genetic alphabets (dzA > G > C > T). In this condition, the relative peak values of these four Gaussian curves (Rel. G) were limited within 0.6-0.8 for T, 0.8-1.0 for C and 1.0 < for dzA where the peak value of the Gaussian curve of G was defined as 1.0. (3) Baseline current was determined by the peak current value of the lowest fitting curve. (4) Raw current values of all data points in the current trace of interest were converted to current intensity (current intensity = raw current value – baseline current). (5) Current intensity of each data point was normalized by the peak current value of the Gaussian

fitting curve assigned to G. (6) The genetic alphabets were assigned to each plateau of the current trace as probable cases based on the probability density functions (equivalent to the Gaussian fittings).

8. Simulation study of accuracy of base distinguishment (Figure S6)

The number of signal data points (N) that is necessary to distinguish a pair of nucleobases (binary classification) was estimated by the following procedures. (1) The logarithmic Gaussian fitting curves at higher conductance values were extracted from the conductance histograms (Figure 4c) as probability density functions of molecular conductance of nucleobases (dC, dG, dA and dzdA). (2) For each of two nucleobases of interest (dA/dC, dA/dG, dzdA/dC or dzdA/dG nucleobases), 10000 of random conductance values (20000 in total) that follow their probability density functions were generated. (3) The probabilities of correct base assignment ($p(x)$) were calculated based on probability density functions for all the randomly generated conductance values. (4) The average accuracy of base assignment ($\bar{p}(x)$) was calculated by averaging all the $p(x)$ values. (5) The estimated number of signal data points (N) that is required to distinguish two nucleobases of interest at given accuracy (P) was plotted based on the following equation[3].

$$P = \exp\{-(\frac{\bar{p}(x)}{1-\bar{p}(x)})^{-N}\}$$

This simulation study was conducted using three patterns of probability density function with different standard deviations (σ = experimental values, 10% and 8% of G_H, respectively). The latter two cases represent virtual cases assuming the future technological advancement enabling the regulation of DNA orientation and motion to reduce distribution of molecular conductance. The experimental values of σ were within 10–13% of G_H in the current system.

9. Enzymatic incorporation of dzdA into DNA samples (Figure S5)

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'- CACTAGGACGGCAGAGGAACA**T***TCTCCCTATAGTGAGTCGTATTA-3'

(*Bold "T": A starting point of extension where dA or dzdA is enzymatically incorporated)

For PCR amplification of a hTNFα coding region (NCBI Reference Sequence: #NM_000594.3), forward and reverse primers were purchased from Fasmac. A plasmid vector coding human TNFα cDNA (pBluescript SK Ⅱ hTNFα) was obtained from RIKEN. The sequences of primers are as follows.

Forward primer: 5'-ATGAGCACTGAAAGCATGATCCG-3'

Reverse primer : 5'-TCACAGGGCAATGATCCC-3'

Primer extension

The primer (0.4 μM) and template (0.4 μM) were mixed in a 10 μL 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 ºC for 2 minutes and then annealed by cooling to 54 ºC at the rate of –0.1 ºC/sec. Extension reaction was started by adding a 10 μL of 400 μM dNTP mix containing dATP (Sigma-Aldrich) or dzdATP (TriLink BioTechnologies) as a genetic alphabet "A" and the reaction mixture was kept at 54 ºC. After a defined time, reaction was stopped by adding a 20 μL of loading buffer (7 M urea and 20 mM of EDTA in TBE buffer). A 10 μL 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 (FLA-7000, FUJIFILM).

PCR amplification of human TNFα gene

Human TNFα gene (702 bp) was amplified with PCR using pBluescript SKⅡ hTNFα as a template. PCR was performed in 100 μL total reaction volume with primers (1 μM), dNTP mix containing dATP or dzdATP (200 μM), $MgSO_4$ (1.5 mM), 1× PCR buffer for KOD –Plus- Neo (TOYOBO), and 1 unit of KOD –Plus- Neo (TOYOBO). The reaction mixture was first heated at 95 ºC for 2 minutes and followed by 30 cycles of denaturing (95 ºC for 30 seconds), annealing (50 ºC for 30 seconds) and elongation (72 ºC for 60 seconds).

Transformation and cloning for Sanger sequencing of the amplified gene

The PCR products were purified by FastGene PCR Extraction kit following the manufacture's protocol (NIPPON Genetics). The purified PCR products were mixed with 200 μM of dNTP mix, 1.5 mM $MgSO_4$, 1× KOD buffer and 1×A-attachment mix (TOYOBO) to attach canonical dA at 3' end of the amplified DNA. After A-attachment reaction, a 3 μL of resulting mixture was mixed with a 1 μL of pTA2 vector solution (50 ng/μL, TOYOBO), a 5 μL of 2×ligation buffer (TOYOBO) and a 1 μL of T4 ligase (TOYOBO) and then incubated for 30 minutes at room temperature. The ligation products were used for standard *E. coli* transformation and cloning. The sequences of inserted gene from five colonies were verified with Sanger sequencing (Eurofins Genomics) using the following primers.

Sequence_1 (Forward): 5'-TAATACGACTCACTATAGGG-3'

Sequence_2 (Reverse) : 5'-CAGGAAACAGCTATGACCATG-3'

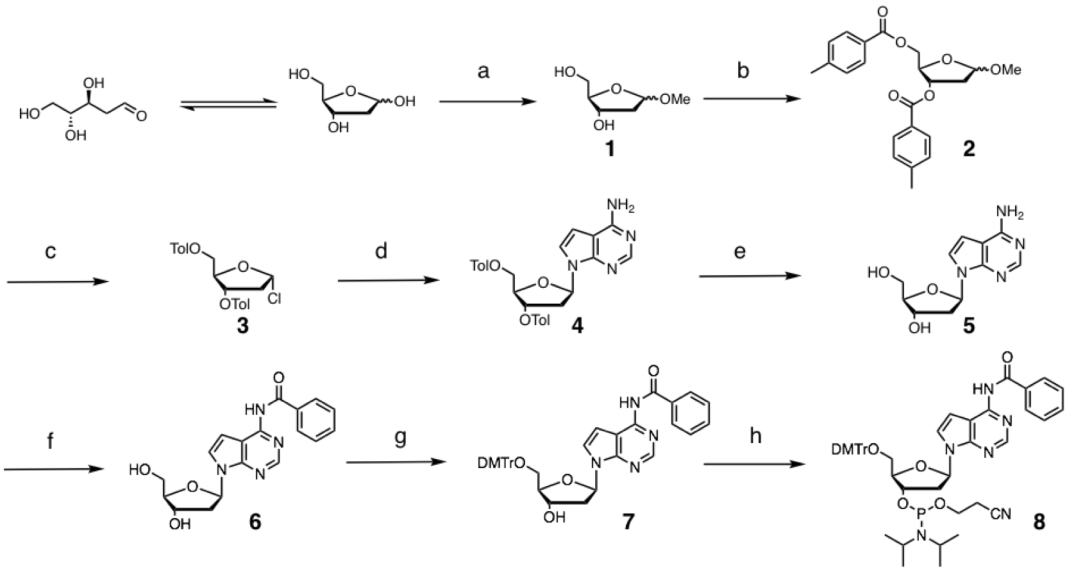
10. Synthesis and purification of oligonucleotides

Synthesis of nucleotides containing ddR was performed using automatic DNA synthesizer (M-2-MX, Nihon Techno Service) on a 1.0 μmole scale (500 Å CPG beads as solid supports). Each phosphoramidite was dissolved in anhydrous acetonitrile (ACN) to give 0.1 M solution and it was pushed into the CPG column simultaneously with equal volume of 0.25 M 4,5-dicyanoimidazole in anhydrous ACN as an activator for phosphoramidite coupling. Coupling time of synthesized phosphoramidite was typically 3 minutes, almost doubled compared to standard coupling time for commercially available phosphoramidite to ensure sufficient coupling yield. After coupling of phosphoramidite, standard capping and oxidation steps were performed. After all coupling steps were completed, terminal 4,4'-dimethoxytrityl (DMTr) protecting groups were cleaved using a standard DMTr off procedure. The resulting oligonucleotides on CPG beads were cleaved by treatment with 28% NH_3 aqueous solution at room temperature for 1 hour. Then the resulting solution was heated at 55 ºC for 16 hours to remove protecting groups from oligonucleotide strands. After cooling down, the oligonucleotide solution was freeze-dried and purified by reverse phase HPLC eluting with 100 mM TEAA (pH 7.0)/ACN (ACN: 2% for 5 min, linear gradient of ACN: 2% – 25% for 23 minutes). Eluted oligonucleotides were monitored by UV absorbance at 260 nm and the fractions containing target products were lyophilized to remove the solvent. The resulting dry oligonucleotides were dissolved in deionized water and characterized by MALDI-TOF MS (reflector positive mode) (Table S3).

11. Chemical synthesis

NMR spectra were recorded on an ECS400 spectrometer (JEOL). Tetramethylsilane (0.00 ppm) in $CDCl_3$, a residual peak of DMSO (2.50 ppm) in DMSO-d_6 were used as internal standards for ^{1}H NMR. $CDCl_3$ (77.2 ppm), DMSO-d_6 (39.5 ppm) and MeOD (49.0 ppm) were used as internal standards for ^{13}C NMR. 85% H_3PO_4 (0.00 ppm) was used as an external standard for ^{31}P NMR. ESI mass spectra (ESI-MS) were measured using a micrOTOF II (Bruker).

Synthesis of 7-deaza dA phosphoramidite



Scheme S1 a. 1–2% HCl in MeOH, rt, 20 min, quant.; **b.** p-toluoyl chloride (2.2 eq.), 4-dimethylaminopyridine (0.10 eq.) in pyridine, rt, 16 h, 92%; **c.** 4M HCl in AcOEt (6.0 eq.), 0 ºC, 1 h, 70%; **d.** KOH (2.2 eq.), tris[2-(2-methoxyethoxy)ethyl]amine (0.06 eq.), 4-amino-7H-pyrrolo[2,3-d]pyrimidine (1.0 eq.) in acetonitrile, rt., 30 min, 43%; **e.** 6M NH_3 in MeOH/H_2O (160 eq.), 60 ºC, overnight, 92%; **f.** i) trimethylsilyl chloride (2.4 eq.) in pyridine, rt., 1 h, ii) benzoyl chloride (1.2 eq.), rt., 1 h, iii) 28% NH_3 aqueous solution, rt., 2h, 95%; **g.** 4,4'-dimethoxytrityl chloride (1.2 eq.) in pyridine, rt., overnight, 72%; **h.** 2-cyanoethyl diisopropylchlorophosphoramidite (1.2 eq.), diisopropylethylamine (3.0 eq.) in dichloromethane, rt., 4 h, 84%.

7-Deaza dA phosphoramidite was synthesized as shown in Scheme S1. The synthesis of 7-deaza dA (**5**)[4] and phosphoramidite compound **8**[5] were conducted according to the literature.

Characterization of compounds 3, 5 and 8

2'-Deoxy-3',5'-di-*O*-*p*-toluoyl-α-D-ribofuranosyl chloride (3)

White powder. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.99 (d, J = 8.2 Hz, 2H), 7.90 (d, J = 8.2 Hz, 2H)), 7.39-7.21 (m, 4H), 6.47 (d, J = 5.0 Hz, 1H), 5.58-5.53 (m, 1H), 4.88-4.83 (m, 1H), 4.71-4.56 (m, 2H), 2.91-2.71 (m, 2H), 2.42 (s, 3H), 2.41 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 166.5, 166.2, 144.4, 144.2, 130.1, 129.8, 129.3(8), 129.3(6), 126.9, 126.8, 95.5, 84.8, 73.7, 63.6, 44.7, 21.8(8), 21.8(4).

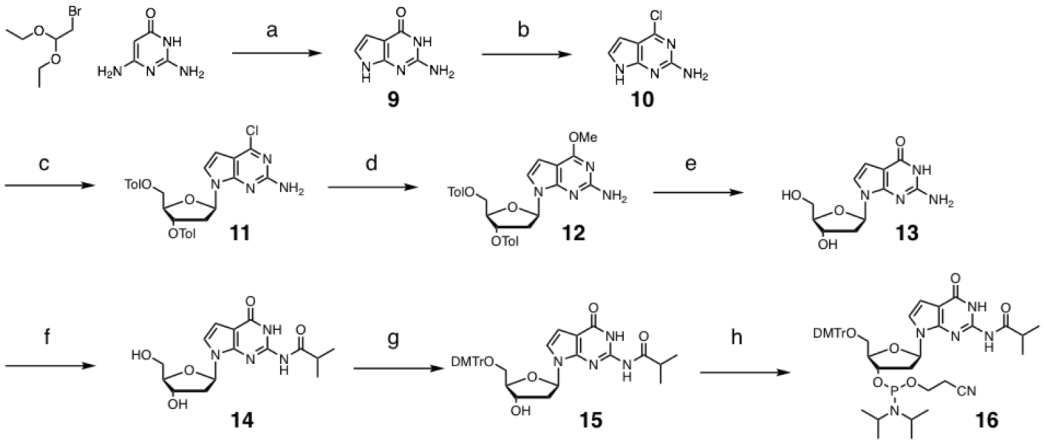
7-Deaza-2'-deoxyadenosine (7-deaza dA, 5)

White form. ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.04 (s, 1H), 7.33 (d, *J* = 3.7 Hz, 1H), 7.01 (br s, 2H), 6.57 (d, *J* = 3.7 Hz, 1H), 6.47 (dd, *J* = 8.2, 5.9 Hz, 1H), 5.23 (d, *J* = 4.1 Hz, 1H), 5.14 (t, *J* = 5.5 Hz, 1H), 4.37-4.31 (m, 1H), 3.84-3.79 (m, 1H), 3.60-3.39 (m, 2H), 2.55-2.44 (m, 1H, overlapped with DMSO signal), 2.18-2.10 (m, 1H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 157.5, 151.6, 149.6, 121.6, 102.9, 99.6, 87.3, 83.3, 71.1, 62.1, 39.5 (overlapped with DMSO signal); HRMS (ESI) m/z: $[M+Na]^+$ Calcd for $C_{11}H_{14}N_4NaO_3$ 273.0958; found 273.0945.

6-*N*-Benzoyl-7-deaza-3'-(2-cyanoethyl-*N*,*N*-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)adenosine (8)

Obtained as diastereomer mixture. White form. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.70 (s, 1H), 8.54 (s, 1H), 8.01-7.95 (m, 2H), 7.64-7.58 (m, 1H), 7.56-7.50 (m, 2H), 7.47-7.17 (m, 9H), 7.02-04 (m, 1H), 6.86-6.77 (m, 5H), 4.78-4.68 (m, 1H), 4.29-4.23 (m, 1H), 3.97-3.54 (m, 10H), 3.45-3.28 (m, 2H), 2.71-2.43 (m, 4H), 1.23-1.08 (m, 12H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 165.3(4), 165.2(9), 158.6, 153.3, 150.6, 150.2, 144.7, 135.8(8), 135.8(5), 133.7, 132.8, 130.2, 129.0, 128.4, 128.3, 128.0, 127.8, 127.0(2), 126.9(9) 123.8, 117.6(3), 117.5(6), 113.2, 109.0, 104.8, 86.5, 85.3(3), 85.2(9), 85.1(5), 85.0(9), 83.6, 74.4, 74.2, 73.7, 73.6, 63.7, 63.5, 58.5(8), 58.5(3), 58.4(0), 58.3(4), 55.4, 43.4(6), 43.4(1), 43.3(3), 43.2(8), 39.9, 24.8(0), 24.7(7), 24.7(0), 24.6(3), 24.6(1), 20.5(3), 20.4(7), 20.3(6), 20.2(9); ^{31}P NMR (160 MHz, $CDCl_3$) δ 149.4, 149.1; HRMS (ESI) m/z: $[M-H]^-$ Calcd for $C_{48}H_{52}N_6O_7P$ 855.3641; found 855.3647.

Synthesis of 7-deaza dG phosphoramidite



Scheme S2 a. NaOAc (2.2 eq.), HCl (0.25 eq.) in water, 80 ºC, 20 min, 56%; **b.** pyridine (1.0 eq.) in $POCl_3$ (10 eq.), microwave at 120 ºC, 30 min, 49 %; **c**. compound 3 (1.1 eq.), NaH (1.1 eq.) in acetonitrile, rt, 79%; **d**. NaOMe (7.0 eq.) in MeOH, reflux, 4 h, 67 %.; **e**. 2M NaOH (7.0 eq.) in water, reflux, 6 h, quant.; **f.** i) trimethylsilyl chloride (10 eq.) in pyridine, rt., 2 h, ii) isobutyryl anhydride (5.0 eq.), rt., 4 h, iii) 28% NH_3 aqueous solution, rt., 1h, 81%; **g**. 4,4'-dimethoxytrityl chloride (1.2 eq.) in pyridine, rt., overnight, 76%; **h.** 2-cyanoethyl diisopropylchlorophosphoramidite (2.0 eq.), diisopropylethylamine (3.0 eq.) in dichloromethane, rt., 3 h, 64%.

7-Deaza dG phosphoramidite was synthesized as shown in Scheme S2. The synthesis of 7-deaza guanine (**9**)[6], 2-amino-6-chloro-7-deazapurine (**10**)[7], 7-deaza dG (**13**)[8] and synthesis of phosphoramidite compound **16**[9] were conducted according to the literature.

Characterization for compounds 13 and 16

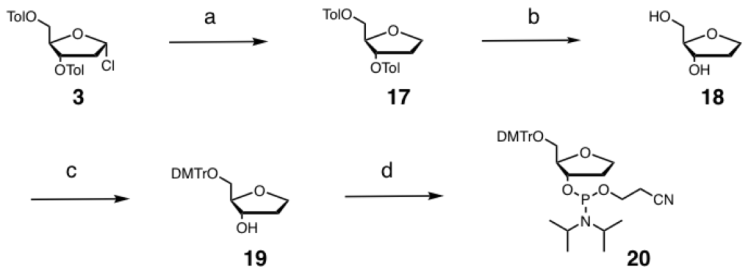
7-Deaza-2'-deoxyguanosine (7-deaza dG, 13)

Off white solid. ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.62 (s, 1H), 6.90 (d, *J* = 3.7 Hz, 1H), 6.51, (br s, 2H), 6.29 (dd, *J* = 8.2, 5.9 Hz, 1H), 6.25 (d, *J* = 3.7 Hz, 1H), 5.28 (br s, 1H), 4.94 (br s, 1H), 4.31-4.26 (m, 1H), 3.78-3.72 (m, 1H), 3.53-3.44 (m, 2H), 2.36-2.26 (m, 1H), 2.12-2.03 (m, 1H); ^{13}C NMR (100 MHz, methanole-d_4) δ 162.0, 153.8, 152.4, 119,8, 103.3, 102.0, 88.6, 85.6, 73.0, 63.8, 41.1; HRMS (ESI) m/z: $[M+Na]^+$ Calcd for $C_{11}H_{14}N_4NaO_4$ 289.0907; found 289.0907.

7-Deaza-2-*N*-isobutyryl-3'-(2-cyanoethyl-*N,N*-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)guanosine (16)

Obtained as diastereomer mixture. White form. ^{1}H NMR (400 MHz, $CDCl_3$): δ 11.73 (br s, 1H), 8.53(br s, 0.5H), 8.20 (br s, 0.5H), 7.48-7.41 (m, 2H), 7.36-7.29 (m, 4H), 7.28-7.17 (m, 3H), 6.90-6.88 (m, 1H), 6.82-6.76 (m, 4H), 6.63-6.60 (m, 1H), 6.43-6.35 (m, 1H), 4.74-4.58 (m, 1H), 4.28-4.17 (m, 1H), 3.90-3.52 (m, 4H), 3.78 (s, 6H), 3.41-3.22 (m, 2H), 2.78-2.26 (m, 5H), 1.22-1.08 (m, 18H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 178.6, 178.3, 158.7, 157.9(4), 157.8(7), 147.5(8), 147.5(1), 146.1, 146.0, 144.8, 144.7, 136.0, 135.9(1), 135.8(7), 135.8(3), 130.2(3), 130.1(6), 128.4, 128.3, 128.0, 127.0, 119.9, 119.3, 117.8, 117.7, 113.2(7), 113.2(3), 106.1, 105.7, 103.9, 86.5, 86.4, 85.3, 84.3(9), 84.3(6), 74.6, 74.5, 73.9, 73.8, 63.9(5), 63.8(7), 58.4, 58.2, 55.4, 43.4(1), 43.3(5), 43.2(8), 43.2(3), 39.9(8), 39.9(4), 39.5(4), 39.4(9), 36.4, 36.2, 24.7(4), 24.7(0), 24.6(7), 20.5(0), 20.4(3), 20.3(6), 19.0(9), 19.0(4), 19.0(0), 18.9(7); ^{31}P NMR (160 MHz, $CDCl_3$) δ 148.6, 148.2; HRMS (ESI) m/z: $[M+Na]^+$ Calcd for $C_{45}H_{55}N_6NaO_8P$ 861.3711; found 861.3694.

Synthesis of 1',2'-dideoxy-D-ribofuranose (ddR) and its phosphoramidite compound



Scheme S3. a. Tributyltin (1.0 eq.), 2,2'-azobisisobutyronitrile (0.10 eq.) in toluene, 60 ºC, 2.5 h, 100 %; **b**. NH_3 in MeOH/Water, 16 h, 65 ºC, 65 %. **c**. 4,4'-dimethoxytrityl chloride (1.2 eq.) in pyridine, rt., 14 h, 32%; **d**. 2-cyanoethyl diisopropylchlorophosphoramidite (1.2 eq.), diisopropylethylamine (3.0 eq.) in dichloromethane, rt., 2 h, 97%.

Starting from compound **3**, 1',2'-dideoxy-D-ribofuranos (**18**)[10] and phosphoramidite compound **20**[11] was synthesized according to the literature (Scheme S3).

Characterization of Compound 18 and 20

1,2-Dideoxy-D-ribofuranose (18)

Colorless syrup. ^{1}H NMR (400 MHz, $CDCl_3$): δ 4.35-4.29 (m, 1H), 4.04-3.97 (m, 2H), 3.82 (q, *J* = 4.1 Hz, 1H), 3.77-3.71 (m, 1H), 3.67-3.60 (m, 1H), 2.19-2.09 (m, 1H), 1.97-1.89 (m, 1H), 1.89-1.83 (br, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 86.4, 73.0, 67.2, 62.9, 35.3; HRMS (ESI) m/z: $[M–H]^-$ Calcd for $C_5H_{11}O_3$ 117.0557; found 117.0557.

3-(2-Cyanoethyl-*N,N*-diisopropylphosphoramidyl)-1,2-dideoxy-5-*O*-(4,4'-dimethoxytrityl)-D-ribofuranose (20)

Obtained as diastereomer mixture. Colorless syrup. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.47-7.42 (m, 2H), 7.36-7.31 (m, 4H), 7.30-7.24 (m, 2H), 7.23-7.16 (m, 1H), 6.85-6.79 (m, 4H), 4.44-4.36 (m, 1H), 4.08-4.00 (m, 2H), 3.99-3.92 (m, 1H), 3.85-3.72 (m, 1H), 3.79 (s, 3H), 3.78 (s, 3H), 3.71-3.49 (m, 3H), 3.20-3.06 (m, 2H), 2.61-2.57 (m, 1H), 2.46-2.42 (m, 1H), 2.17-1.93 (m, 2H), 1.22-1.05 (m, 12H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 158.5, 145.1, 136.2(9), 136.2(5), 130.2(6), 130.2(3) 128.4(0), 128.3(5), 127.9, 126.8(5), 126.8(1), 117.7(0), 117.6(6), 113.2, 86.2, 85.0(5), 85.0(1) 84.8(5), 84.7(9), 76.0, 75.8, 75.6, 75.4, 67.6(1), 67.5(8), 64.4, 64.3, 58.5(1), 58.4(3), 58.3(3), 58.2(3), 55.3, 44.3(8), 44.3(3), 44.2(6), 44.2(1), 34.5(6), 34.5(2), 34.4(6), 34.4(3), 24.7(6), 24.6(9), 24.6(1), 24.5(4), 20.5, 20.4, 20.3(4), 20.2(8); ^{31}P NMR (160 MHz, $CDCl_3$) δ 148.6, 148.3; HRMS (ESI) m/z: $[M+Na]^+$ Calcd for $C_{35}H_{45}N_2NaO_6P$ 643.2907; found 643.2904.

12. References

1) Agrait, N.; Yeyati, A. L.; van Ruitenbeek, J. M. Quantum Properties of Atomic-Sized Conductors. *Phys. Rep.* **2003**, *377*, 81–279.

2) Tsutsui, M.; Taniguchi, M.; and Kawai, T. Fabrication of 0.5 nm Electrode Gaps Using Self-Breaking Technique. *Appl. Phys. Lett.* **2008**, *93*, 163115.

3) Churchill, G. A.; Waterman, M. S. The Accuracy of DNA Sequences: Estimating Sequence Quality. *Genomics* **1992**, *14***,** 89–98.

4) Maiti, M.; Michielssens, S.; Dyubankova, N.; Maiti, M.; Lescrinier, E.; Ceulemans, A.; Herdewijn, P. Influence of the Nucleobase and Anchimeric Assistance of the Carboxyl Acid Groups in the Hydrolysis of Amino Acid Nucleoside Phosphoramidates. *Chem. Eur. J.* **2012**, *18*, 857–868.

5) Seela, F.; Kehne, A. Palindromic Octa- and Dodecanucleotides Containing 2'-Deoxytubercidin: Synthesis, Hairpin Formation, and Recognition by the Endodeoxyribonuclease *Eco*RI. *Biochemistry* **1987**, *26*, 2232–2238.

6) Bruckl, T.; Thoma, I.; Wagner, A. J.; Knochel, P.; Carell, T. Efficient Synthesis of Deazaguanosine-Derived tRNA Nucleosides $PreQ_0$, $PreQ_1$, and Archaeosine Using the Turbo-Grignard Method. *Eur. J. Org. Chem*. **2010**, *34*, 6517–6519.

7) Klepper, F.; Jahn, E. M.; Hickmann, V.; Carell, T. Synthesis of the Transfer-RNA Nucleoside Queuosine by Using a Chiral Allyl Azide Intermediate. *Angew. Chem. Int. Ed.* **2007**, *46*, 2325–2327.

8) Ramasamy, K.; Imamura, N.; Robins, R. K.; Revankar, G. R. A Facile and Improved Synthesis of Tubercidin and Certain Related Pyrrolo[2,3-*d*]pyrimidine Nucleosides by the Stereospecific Sodium Salt Glycosylation Procedure [1]. *J. Heterocyclic Chem*. **1988**, *25*, 1893–1898.

9) Seela, F.; Driller, H. Alternating $d(G\text{-}C)_3$ and $d(C\text{-}G)_3$ Hexanucleotides Containing 7-Deaza-2'-Deoxyguanosine or 8-Aza-7-Deaza-2'-Deoxyguanosine in Place of dG. *Nucl. Acids. Res*. **1989**, *17*, 901–910.

10) Takeshita, M.; Chang, C. N.; Johnson, F.; Will, S.; Grollman, A. P. Oligodeoxynucleotides Containing Synthetic Abasic Sites. *J. Biol. Chem*. **1987**, *262*, 10171–10179.

11) Perrett, A. J.; Dickinson, R. L.; Krpetic, Z.; Brust, M.; Lewis, H.; Eperon, I. C.; Burley, G. A. Conjugation of PEG and Gold Nanoparticles to Increase the Accessibility and Valency of Tethered RNA Splicing Enhancers. *Chem. Sci.* **2013**, *4*, 257–265.