Nanoscale

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d0nr09031j

Nanopore decoding for a Hamiltonian path problem†

Sotaro Takiguchi and Ryuji Kawano *

DNA computing has attracted attention as a tool for solving mathematical problems due to the potential for massive parallelism with low energy consumption. However, decoding the output information to a human-recognizable signal is generally time-consuming owing to the requirement for multiple steps of biological operations. Here, we describe simple and rapid decoding of the DNA-computed output for a directed Hamiltonian path problem (HPP) using nanopore technology. In this approach, the output DNA duplex undergoes unzipping whilst passing through an α-hemolysin nanopore, with information electrically decoded as the unzipping time of the hybridized strands. As a proof of concept, we demonstrate nanopore decoding of the HPP of a small graph encoded in DNA. Our results show the feasibility of nanopore measurement as a rapid and label-free decoding method for mathematical DNA computation using parallel self-assembly.

Received 22nd December 2020,
Accepted 1st March 2021

DOI: 10.1039/d0nr09031j

rsc.li/nanoscale

Introduction

DNA plays an essential role in nature of memory storage, with information transferred and/or copied *via* chemical and enzymatic reactions. This biological system has attracted the attention of computer scientists, who wish to apply computational methods using wet-lab experimentation. In 1994, the computer scientist Leonard M. Adleman launched the research field of DNA computing with a demonstration that used DNA molecules to solve a directed Hamiltonian path problem (HPP).[1] The solution to this problem involves finding a route among several nodes on a graph, such that each node is visited exactly once, generally known as a traveling salesman problem (finding a sales route between several cities, such that each city is visited only once). This type of combinatorial problem, which is mathematically classified as NP-complete, requires numerous computations. Adleman's proposed methodology addresses this complexity by exploiting the massive parallelism of DNA self-assembly, resulting in the generation of a DNA library encoding all potential routes through the graph. Based on Adleman's methodology, Hsieh *et al.* had also demonstrated the DNA-based graph encoding computation, such as the HPP, from a theoretical perspective.[2] Following this groundbreaking demonstration, several mathematical models were implemented using DNA or RNA molecules,[3–9] and the research field of DNA self-assembly technology has recently expanded from mathematical computation to the development of a variety of nanoscale structures and devices.[51,52] Although this Adleman–Lipton model has been established in molecular computing owing to its massive parallelism and low energy consumption, the output information is encoded in nucleic acid molecules which then requires decoding into a human-recognizable signal. The conventional decoding approach is time-consuming since it requires multiple steps of biological operation, including repetitive magnetic bead experiments and graduated PCR.[1] Answer determination has also been accomplished by DNA sequencing[7] and multiple-gradient gel electrophoresis;[10] these methods are also time-consuming. As an alternative rapid decoding approach, DNA surface computing that utilizes DNA immobilized on a surface with simultaneous computation and purification,[11] and fluorescent probe-mediated RT-PCR[12,13] have been proposed. Although these approaches were relatively readily available, the decoding time was not significantly shortened and both methods require the use of direct labeling of DNAs and fluorescence detection.

Nanopore technology is a promising method for the rapid and label-free detection of target molecules,[14–19,53] and a data analysis method combining artificial intelligence has recently enhanced the research field.[20,21] When a molecule passes through a nanopore under an applied voltage, it can be recognized by ion current blocking. In particular, α-hemolysin (αHL), a pore-forming toxin from *Staphylococcus aureus*,[22] is conventionally used as a biological nanopore for detecting oligonucleotides based on the size-matching between the pore and single-stranded DNA (ssDNA).[23–27] A ssDNA can pass through the nanopore, whereas a double-stranded DNA (dsDNA) cannot pass and clogs the pore vestibule owing to the

Department of Biotechnology and Life Science, Tokyo University of Agriculture and Technology, Tokyo, 184-8588, Japan. E-mail: rjkawano@cc.tuat.ac.jp

†Electronic supplementary information (ESI) available. See DOI: 10.1039/d0nr09031j

size mismatch. Because the results of DNA computation are output as DNA or RNA molecules,[28–32] DNA computing and αHL nanopore technology have been integrated.[33–38] Firstly, the rapid and label-free decoding of oligonucleotides in logic gate-type DNA computation was proposed.[33] In this system, input DNAs were injected into a droplet, where the logic operation was performed, and the result was encoded in DNA as a structural change. The output molecules were translocated to another droplet through a nanopore, which was monitored electrically, and decoded. A logic network[34] and a more complex logic operation including enzymatic reactions[36] were also constructed using this droplet system. The significance of these studies is the definition of the output information of "0" and "1" in the case of nanopore measurement. The output information of 0 or 1 in this logic operation was defined according to whether ssDNA or RNA was translocated through a nanopore. Following these studies, Hiratani *et al.* applied these methodologies to microRNA (miRNA) pattern recognition that can diagnose cancer from bodily fluid as a liquid biopsy.[37,38] The expression pattern of two overexpressed miRNAs was described as an AND logic operation, resulting in the formation of a four-way junction with diagnostic DNAs. Nanopore measurement was able to discriminate the AND operations; (0, 0), (0, 1), (1, 0), and (1, 1), by analyzing the blocking duration and unzipping time, using a bootstrapping method. Bootstrapping of the data offered a significant improvement in distinguishing the patterns in cases where the raw unzipping time series data had been unable to distinguish any difference.[38,39]

Based on these nanopore decoding technologies, we here attempt to decode a complex mathematical DNA computation with parallel self-assembly, including the analysis of the unzipping time by bootstrapping. As a proof of concept, we demonstrated nanopore decoding using a simple graph, which has 5 nodes and 10 paths, encoding the HPP (Fig. 1a). In this approach, the output duplex formed a specific structure and passed through the nanopore with unzipping of the hybridized strands. The information encoded in the output DNA was electrically decoded as the unzipping time by the αHL nanopore.

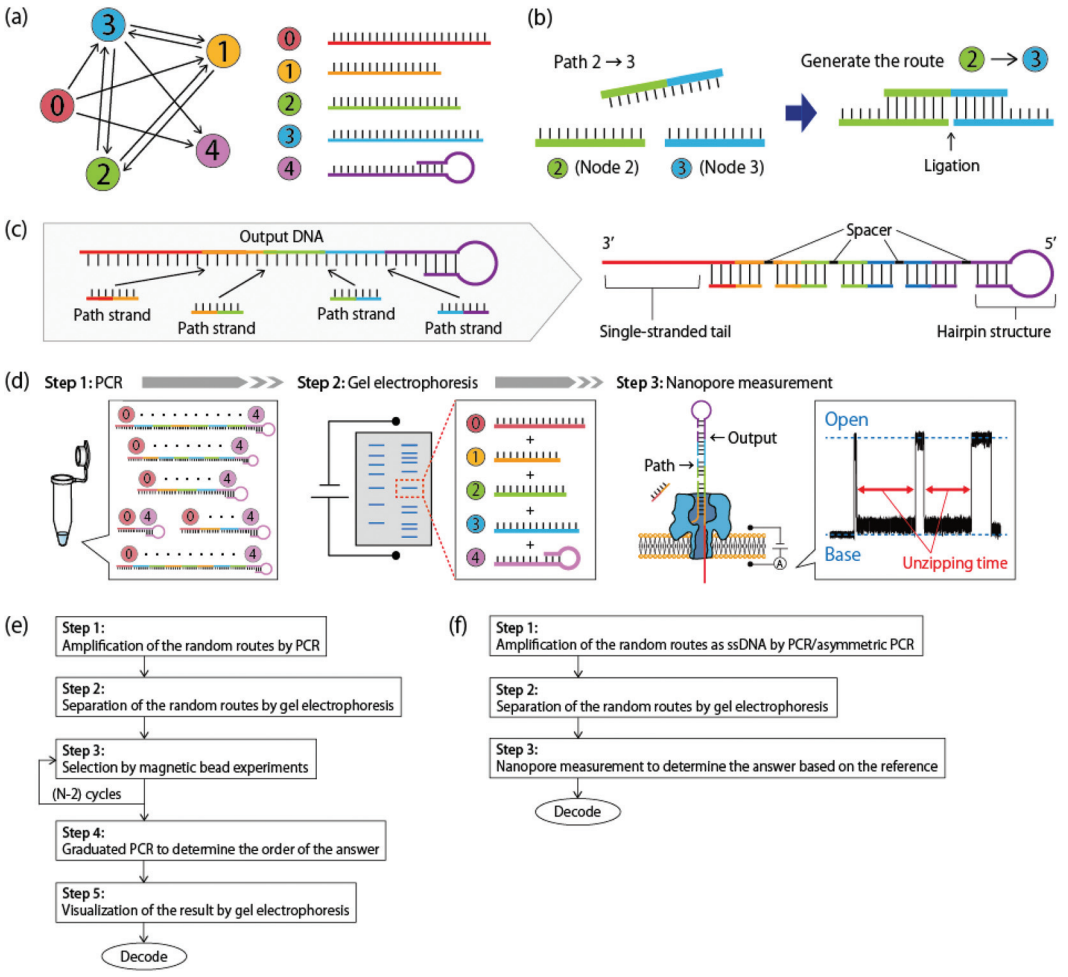

Fig. 1 (a) A directed graph which has a Hamiltonian path, 0 → 1, 1 → 2, 2 → 3, 3 → 4, when start node = 0 and end node = 4. (b) The illustration of the hybridization between node strand and path strand. (c) The illustration of the duplex structure consisting of the output DNA and path strands. (d) The schematic illustration of the nanopore decoding method using an α-hemolysin pore. (e) The flowchart of the conventional method. *N* is the number of the node. (f) The flowchart of nanopore decoding.

Results

Design of DNA sequences as nodes and paths in the graph

To implement DNA-based parallel computation, each "node" in the graph should be associated with a specific sequence of oligonucleotides. Each "path" in the graph should also be associated with oligonucleotides, which consist of a sequence with complementarity of the node strands, and hence facilitating the connection of two nodes by hybridization as shown in Fig. 1b. These DNA molecules hybridize through parallel self-assembly that generates many random routes through the graph by mixing, hybridizing, and fixing with the ligation reaction. To perform these procedures, we designed the DNA sequences of the nodes and paths for nanopore decoding by the following steps:

(1) We need to design 5 (0 to 4) nodes and 10 paths using oligonucleotides in the graph. The length of oligonucleotides in each node randomly differs so that the output DNA has a specific length that can be identified by gel electrophoresis.

(2) The sequence of node 0 has no secondary structure at the 24 nt of the 3′ terminus. This region plays the role of the single-stranded tail of the duplex (Fig. 1c), to enter the αHL pore from the 3′ direction. The length of 24 nt is sufficiently long to insert from the *cis* to the *trans* side of the αHL nanopore and to remain there.[40]

(3) The sequence of node 4 is designed to form a hairpin structure at the 5′ terminus (Fig. 1c). This hairpin structure restricts the translocating direction, resulting in the output DNA only entering the nanopore from the 3′ terminus side.

(4) The 2 nt of cytosines (–CC–) is inserted as the spacer between each complementary region of the output DNA (Fig. 1c) to induce a tiny difference in the Gibbs free energy (ΔG) among the same length duplexes. To that end, nanopore measurement can recognize the different orders of nodes encoded in the same-length DNA.

(5) In order to prevent unintended hybridization and secondary structure formation, the sequence design is carefully checked using thermodynamic simulation (NUPACK).

Regarding step (4), the number of spacer nucleotides was optimized by using NUPACK simulation. As shown in Fig. S2,† simulations were conducted in the cases that the number of spacer nucleotides was 0, 1, 2 and 3 nt. As listed in Table S1,† the ΔG of the output duplex increased with the increasing of the number of spacer nucleotides, with a 2 nt spacer seen to be sufficient to stabilize the structure. Such an output duplex enters and clogs the constricted region of the nanopore, resulting in the reduction of the ion current flow. Under an applied voltage, the duplex undergoes unzipping whilst passing through the nanopore, with the information electrically decoded as the unzipping time of the hybridized strands (Fig. 1d).

The designed sequences as nodes in this study are listed in Table 1, and the free energies of the 6 kinds of routes that have the same length with the answer are calculated by the NUPACK simulation and listed in Table 2. These 6 kinds of routes satisfy the requirement of the Hamiltonian path problem, and 01234 is the order of answer in the graph. Next, we checked the implementation of this problem using the conventional method.

Table 1 DNA sequences as nodes in the graph

Node	DNA sequence
0	5′ TGGTAAACCTCTGTCACCCCTCTTTCGTCG 3′
1	5′ GCATCCGCCTAATAC 3′
2	5′ GAGGTCGGCCCGCTAATCAGGACT 3′
3	5′ CTACGAATTCCGGCTCAATACTCACGT 3′
4	5′ GTAGAACTTTTGTTCTACCCATATAGTCGAGGTAACGC 3′

Table 2 Gibbs free energies of the duplex of 6 types of routes according to NUPACK

Route	Free energy [kJ mol^{-1}]
0 → 1 → 2 → 3 → 4	−588.1
0 → 1 → 3 → 2 → 4	−589.3
0 → 2 → 1 → 3 → 4	−590.3
0 → 2 → 3 → 1 → 4	−590.3
0 → 3 → 1 → 2 → 4	−590.3
0 → 3 → 2 → 1 → 4	−591.5

Decoding by the conventional method

Procedures of the conventional decoding method. The conventional decoding approach was first performed to confirm the calculation capability of our designed system. The steps of the procedure are as follows (Fig. 1e):

(Conventional method step 1) Amplification of the random routes by PCR

(Conventional method step 2) Separation of the random routes by gel electrophoresis

(Conventional method step 3) Selection by magnetic bead experiments

(Conventional method step 4) Graduated PCR to determine the order of the answer

(Conventional method step 5) Visualization of the result by gel electrophoresis

Implementation of DNA computation and conventional decoding. First, the designed oligonucleotides were mixed to implement DNA-based parallel computation, and the subsequent mixture was checked by gel electrophoresis. As shown in Fig. 2a, the result of DNA computation showed smear-like bands, which reflected the generation of a large number of random routes through the graph. Then, we followed the conventional steps for decoding. In the conventional method, magnetic bead experiments are necessary to purify the routes visiting all nodes exactly once. Owing to the sequence design step (1), we were able to skip this procedure (conventional method step 3). After PCR (conventional method step 1), the output DNA – which encoded a route visiting all nodes exactly once – was extracted from the gel (conventional method step 2) and decoded by graduated PCR (conventional method step 4), which can decode the output information by running four

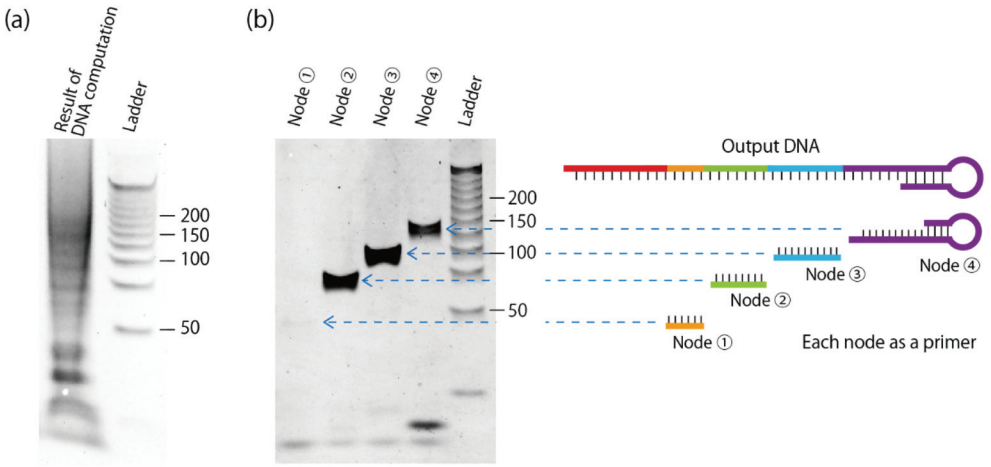

Fig. 2 (a) Result of DNA-based parallel computation. (b) The illustration and the result of graduated PCR which is the conventional decoding approach.

different PCRs. In this method, as shown in Fig. 2b, each node strand is used as a primer, resulting in different lengths of the amplified fragments between node 0 and each node, with the results visualized by gel electrophoresis (conventional method step 5). The order of nodes encoded in the output DNA is determined from the order of the length of each fragment. Fig. 2b shows the result of graduated PCR, indicating that the output DNA encoded the route $0 \rightarrow 1$, $1 \rightarrow 2$, $2 \rightarrow 3$, $3 \rightarrow 4$, which was the answer to the HPP within the graph (the details are shown in Fig. S3†). This designed system took more than six hours to decode the output information by the conventional method even when skipping the time-consuming magnetic bead procedures. With the calculation capability confirmed, we next implemented nanopore decoding using the output molecule of the DNA-based parallel computation.

Decoding by nanopore measurement

Nanopore decoding procedures. The steps of the procedures of nanopore decoding are as follows (Fig. 1f):

(Step 1) Amplification of the random routes as ssDNA by PCR/asymmetric PCR

(Step 2) Separation of the random routes by gel electrophoresis

(Step 3) Nanopore measurement to determine the answer based on the reference

Preparation of the reference data for step 3 in nanopore decoding. In our system, the order of nodes encoded in the output DNA can be decoded as the unzipping time of the output dsDNA using an αHL nanopore (Fig. 1d), and the output DNA which encodes the route visiting all nodes exactly once has already been purified from the random routes through gel electrophoresis. Therefore, the decoding can be performed by comparing the reference data including a series of routes visiting all nodes exactly once. To implement the nanopore decoding, we prepared the reference data as all kinds of routes satisfying the requirement of the HPP, using chemically synthesized barcode-like DNA (bcDNA). We prepared 6 kinds of bcDNA which encoded each route listed in Table 2 and obtained the unzipping time of each duplex by nanopore measurement. Fig. 3a–f shows the typical current-time trace of the unzipping event using each type of bcDNA with its corresponding path strands. The small differences of the ΔG shown in Table 2 were distinguished from the peak values of the histograms (Fig. 3g), indicating that nanopore measurement can discriminate the order of nodes even in the same length duplexes. As shown in Fig. 3h, the unzipping time exponentially increased depending on the ΔG of hybridization simulated by NUPACK. According to this result, nanopore measurement with the bootstrapping method was able to perform precise discrimination of the small differences of ΔG, resulting in the discrimination of the same length routes which composed the reference data. Although the unzipping time of 02134, 02314, and 03124 was almost the same owing to their similar ΔGs, nanopore decoding can be discriminated between the correct answer (01234) and the others. The peak value of the unzipping time of the answer route (01234) was 4688 ms, with one standard deviation from this peak discernible from the peaks of other routes.

Comparison between the output information and reference (steps 1–3). After asymmetric PCR (step 1) and the extraction of the output DNA from the gel (step 2), nanopore measurements (step 3) were conducted using the output DNA and all kinds of path strands (Fig. 4a). Fig. 4b shows a typical current–time trace of the unzipping events, and the event frequency was 0.17 s^{-1} (Fig. S4†). Fig. 4c shows the comparison between the result of nanopore decoding and the reference data. We calculated the overlapping ratio of the histograms between the result of nanopore decoding and each route of the reference (Fig. 4d). As shown in Table 3, the value of the overlapping ratio varied from 4.2% to 73.2%, and a larger overlapping ratio indicates a better correspondence of data. The result of nanopore decoding mostly corresponded with that of the suggested

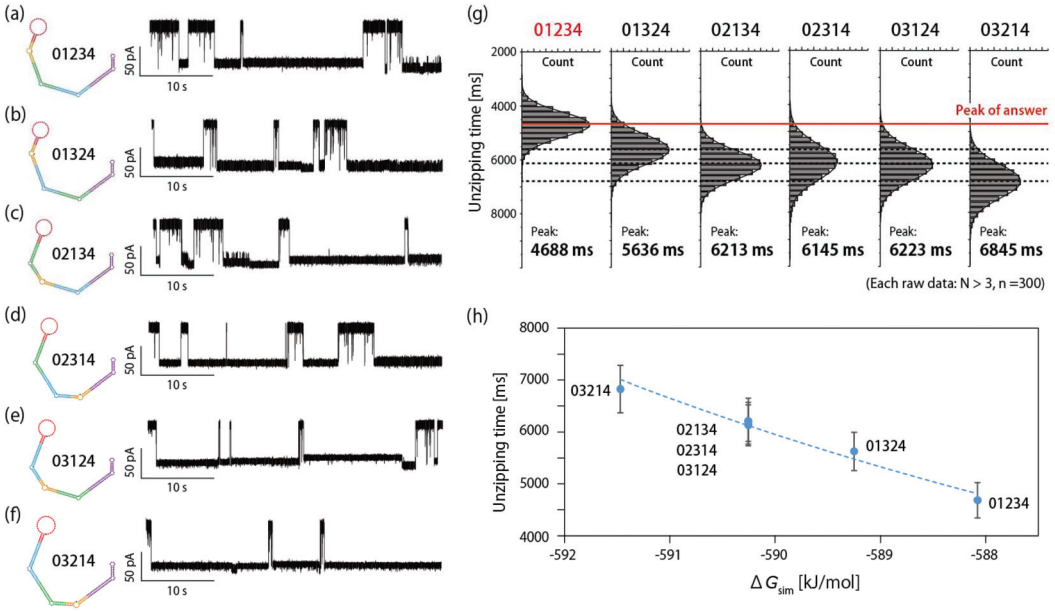

Fig. 3 Result of nanopore measurement with chemically synthesized barcode-like DNA for reference data. Typical current traces of 6 kinds of routes (a) 01234, (b) 01324, (c) 02134, (d) 02314, (e) 03124, and (f) 03214, and each duplex structure was simulated by NUPACK. (g) Histograms of the unzipping time for each route after bootstrapping the data (*N* is the number of the αHL nanopore and *n* is the number of the unzipping signal). (h) Correlation between the unzipping time and the simulated hybridization energy (ΔG_{sim}) calculated using NUPACK.

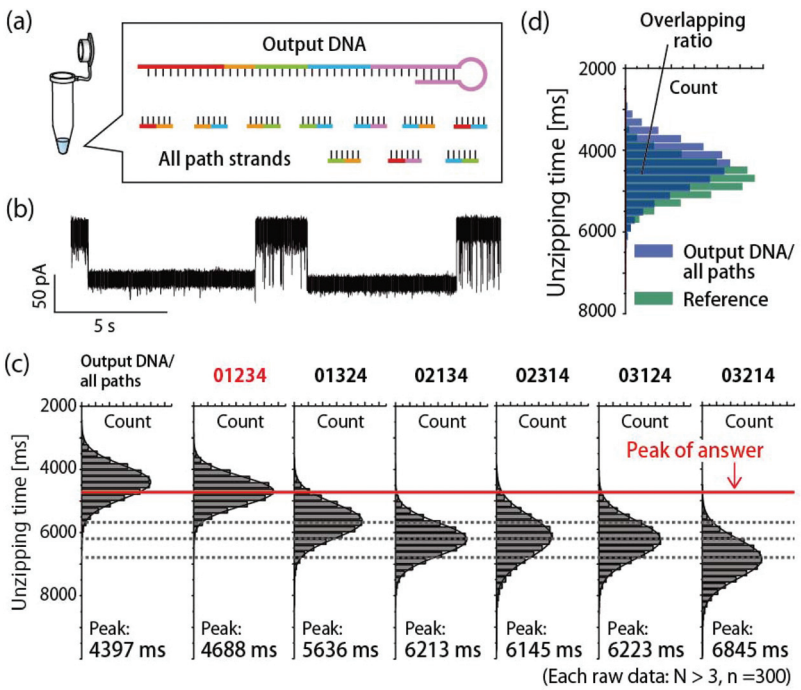

Fig. 4 (a) The illustration of the experimental conditions of the nanopore measurement. (b) Typical current-time trace. (c) The comparison between the result of nanopore decoding and the reference. The bootstrapped data were used to produce the histograms of the unzipping time (*N* is the number of the αHL nanopore and *n* is the number of the unzipping signal). (d) The overlapping ratio of the histograms between nanopore decoding and the reference.

Table 3 The overlapping ratio

Route	Overlapping ratio
$0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4$	73.2%
$0 \rightarrow 1 \rightarrow 3 \rightarrow 2 \rightarrow 4$	26.0%
$0 \rightarrow 2 \rightarrow 1 \rightarrow 3 \rightarrow 4$	10.0%
$0 \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow 4$	14.8%
$0 \rightarrow 3 \rightarrow 1 \rightarrow 2 \rightarrow 4$	12.5%
$0 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 4$	4.2%

answer (01234) with 73.2%, indicating that we were successful in demonstrating the nanopore decoding.

Discussion

The relationship between the unzipping time and ΔG

The relationship between the unzipping time and the ΔG of the duplex is sufficiently significant to predict the unzipping time from ΔG. In the nanopore measurement, the dsDNA blocked the nanopore and translocated with unzipping of the hybridized strands. As shown in Fig. 3h, the unzipping time increased exponentially with the increase of ΔG. The unzipping kinetics can be described as a first-order reaction as follows:[39,41,42]

$$[\text{dsDNA} + \alpha\text{HL}] \rightarrow [\text{ssDNA}] + [\alpha\text{HL}]$$

The dissociation rate constant can be described as follows:

$$k = k_0 \exp(-\Delta G_{\text{sim}}/k_{\text{B}}T) \quad (1)$$

$$k = 1/t. \quad (2)$$

Here, k is the dissociation rate constant, k_0 is the initial rate constant, ΔG_{sim} is the simulated ΔG, k_{B} is the Boltzmann constant used for this single-molecule experiment, T is the temperature, and t is the unzipping time. k should be equal to the inverse of t since $1/t$ is the reaction frequency per unit of time of the single-molecule reaction. As shown in Fig. S5,† the unzipping data of the reference data could be fitted using eqn (1) well, indicating that the inverse value of the unzipping time is the first-order rate constant even in this system.

The feasibility of nanopore decoding

We here discuss the feasibility of nanopore decoding under the condition that an answer is unknown. In this system, there is a strong relationship between the unzipping time and the ΔG of the duplex, indicating that the unzipping time of each route can be predicted from the ΔG of each duplex. Therefore, the route encoded in the output DNA can be obtained from its unzipping time by preparing all of the ΔG of each duplex as the reference. Moreover, as shown in Fig. S6,† nanopore measurements can precisely obtain the information encoded in the output DNA even though all kinds of path strands exist.

We consider the case of a large graph decoded by the nanopore method. The sequences of the oligonucleotides are designed to represent the nodes and paths in the HPP graph, for instance, it has 7 nodes and 14 paths (Fig. S7†). The actual sequence of each node is listed in Table S2.† As shown in Table S3,† there is a small difference of ΔG between the duplexes of the suggested answer and one of the reference routes due to the sequence design. Nanopore measurement would be able to distinguish such a small difference of ΔG with bootstrapping, resulting in decoding of the information and obtaining the order of the answer from the ΔG references.

Comparison with the conventional method

As shown in Fig. 1e and f, nanopore decoding can complete the decoding operations in fewer steps than the conventional approach. Our approach used a specific DNA sequence design to overcome the need for repetitive magnetic bead experiments. Compared with previous approaches towards an alternative to graduated PCR for the determination of the order of nodes,[11–13] nanopore measurement can decode the output information without direct labeling, fluorescence detection, or temperature control. Additionally, the nanopore measurement worked well by just dropping the solution to a microdevice, and the information encoded in DNA molecules was directly converted to the electrical signals. Therefore, nanopore measurement should be less laborious and more timesaving than PCR methods because PCR methods require gel electrophoresis to visualize the result. From the aspect of DNA structure as information storage, PCR requires temperature changing that will break the secondary structure of the output DNA molecules whereas nanopore measurement can directly decode the structural information as unzipping time. In terms of the decoding accuracy, a thermodynamic simulation can precisely predict the structure of the duplex from the DNA sequence and calculate the ΔG, and nanopore measurements obtain precise information from the unzipping time with bootstrapping. As shown in Fig. 4c, we calculated the overlapping ratio of the histograms, with the value sufficiently different to enable the discrimination of the answer routes from the other routes. This distinguishability is potentially comparable to that of the result of gel electrophoresis for visualization of the information encoded in the output DNA.

Although nanopore measurement enables the rapid detection of oligonucleotides, the procedure of nanopore decoding still requires PCR and gel electrophoresis for the following reasons. In the nanopore decoding method, the output information is required to be encoded in ssDNA because the output ssDNA was designed to form a specific duplex with path strands in order to pass through the nanopore with unzipping of the hybridized strands. To this end, the PCR step for the amplification of the random routes as ssDNA is necessary to perform nanopore measurement. In addition, as shown in Fig. 1d, PCR enables the amplification of the routes from node 0 (start node) to node 4 (end node), though in a random order, and gel electrophoresis enables the purification of the routes visiting 5 nodes (all nodes exactly once) from the amplified random routes. At present, it is difficult to employ nanopore technology as an alternative to these steps, resulting in still necessary steps in nanopore decoding. Nanopores can however detect oligonucleotides at a low concentration (~1 fM),[43] so the method of nanopore decoding could potentially perform even more rapid decoding without the need for PCR.

Conclusions

In our designed system, the DNA-computed output formed a specific duplex and underwent unzipping whilst passed through an αHL nanopore, with the information electrically decoded as the unzipping time. The method of nanopore decoding was successfully demonstrated using a small graph encoding the HPP with the statistical processing including a bootstrapping method. Our data showed the feasibility of nanopore decoding for DNA-based parallel computing, with such nanopore decoding applicable to a wide variety of DNA computation. This methodology also has the potential for not only the decoding of DNA storage information[44] but real-life applications, such as miRNA pattern recognition and as a sensor in molecular robots.[45,46]

Experimental

Reagents and chemicals

In this study, we used the following reagents: 1,2-diphytanoyl-*sn-glycero*-3-phosphocholine (DPhPC; Avanti Polar Lipids, Alabaster, AL, USA), *n*-decane (Wako Pure Chemical Industries, Ltd, Osaka, Japan), potassium chloride (KCl; Nacalai Tesque), and 3-morpholinopropane-1-sulfonic acid (MOPS; Nacalai Tesque, Kyoto, Japan). Buffered electrolyte solutions (0.5 M KCl, 5 mM MOPS, pH 7.0) were prepared using ultrapure water, which was obtained from a Milli-Q system (Millipore, Billerica, MA, USA). Wild-type α-hemolysin (αHL; Sigma-Aldrich, St Louis, MO, USA, and List Biological Laboratories, Campbell, CA, USA) was obtained as the monomer polypeptide, isolated from *Staphylococcus aureus* in the form of powder and dissolved at a concentration of 1 mg mL^{-1} in ultrapure water. For use, samples were diluted to the designated concentration using a buffered electrolyte solution and stored at 4 °C. High-performance liquid chromatography grade DNA oligonucleotides were synthesized by Eurofins Genomics Inc. (Tokyo, Japan) and reversed-phase chromatography grade 5′-phosphorylated DNA oligonucleotides were synthesized by FASMAC Co., Ltd (Kanagawa, Japan), respectively, stored at −20 °C. 10× TBE buffer was obtained from Takara Bio Inc. (Shiga, Japan) and was 10-fold diluted for gel electrophoresis. The power supply and an LED transilluminator were obtained from Bio Craft Co., Ltd (Tokyo, Japan) and Optocode Corporation (Tokyo, Japan), respectively. KOD SYBR® qPCR Mix (TOYOBO Co., Ltd, Osaka, Japan) was used for PCR and T4 DNA ligase (Thermo Fisher Scientific Inc., Waltham, MA, USA) was used for the ligation reaction.

DNA design and free energy calculation

To confirm DNA–DNA hybridization, thermodynamic simulations were performed using the NUPACK web-server (California Institute of Technology, http://www.nupack.org/). These analyses were performed at 22 °C using 1 μM DNA in 0.5 M KCl buffered electrolyte solution. The Gibbs free energies (ΔG) were obtained from the NUPACK simulation.

DNA-based parallel computation

Each oligonucleotide (25 pmol) with the 5′-terminal phosphate residues as nodes and paths in the graph, T4 DNA ligase, ligase buffer, and ultrapure water to a total volume of 50 μL was incubated for 1 h at room temperature.

The scheme of nanopore decoding

The flowchart of nanopore decoding is shown in Fig. 1f, and is described in detail here:

Amplifying the random routes as single-stranded DNA. The PCR mixture for amplifying the random routes in a total volume of 20 μL contained 10 μL of KOD SYBR® qPCR Mix, the solution after DNA computation, 0.5 μM each of the forward primer and reverse primer. PCR thermal cycles were carried out using a thermal cycler (astec) to amplify the random routes by the following process: preheat at 98 °C for 1 min; 45 cycles of 98 °C for 10 s, 58 °C for 10 s, 68 °C for 30 s. Then, the asymmetric PCR mixture for amplifying the random routes as single-stranded DNA (ssDNA) in a total volume of 20 μL contained: 10 μL of KOD SYBR® qPCR Mix, the PCR product, and 5 μM of the reverse primer. PCR thermal cycles were carried out using a thermal cycler (astec) to amplify the random routes by the following process: preheat at 98 °C for 1 min; 45 cycles of 98 °C for 10 s, 58 °C for 10 s, 68 °C for 30 s.

Separating the random routes depending on the length. Asymmetric PCR products were separated depending on their length by 10% denaturing polyacrylamide gel electrophoresis (containing 19/1 acrylamide/bis (w/w)) in 1× TBE buffer at a constant power of 7.5 W for 20 min at room temperature. The gel was prepared in our laboratory. After electrophoresis, the gel was stained with diluted SYBR Green II (Takara Bio Inc., Japan) solution for 30 min and visualized under blue LED irradiation; images were obtained using an LED transilluminator (Bio Craft Co., Ltd, Tokyo, Japan). Then the output DNA was extracted from the target band corresponding to ssDNA encoding the route visiting all nodes exactly once using an E.Z.N.A.® Poly-Gel DNA Extraction Kit (Omega Bio-Tek, Inc., Norcross, GA, USA).

Output DNA/path strands hybridization. The nanopore measurement solution contained the output DNA extracted from the gel, and all kinds of path strands (100 nM each in 5 mM MOPS buffer; pH 7.0, containing 0.5 M potassium chloride). The solution was heated at 95 °C for 5 min and then gradually cooled to room temperature.

Preparation of the reference data for nanopore decoding

The nanopore measurement solution contained each chemically synthesized barcode-like DNA and 4 kinds of path strands corresponding to each route (1 μM each in 5 mM MOPS buffer; pH 7.0, containing 0.5 M potassium chloride). The solution was heated at 95 °C for 5 min and then gradually cooled to room temperature.

Preparation of the microdevice

Microdevices were fabricated by machining a 6.0 mm thick, 10 × 10 mm polymethyl methacrylate (PMMA) plate (Mitsubishi

Rayon, Tokyo, Japan) using a computer-aided design and manufacturing with a three-dimensional modeling machine (MM-100, Modia Systems, Japan) as shown in Fig. S1a.† Two wells (2.0 mm diameter and 4.5 mm depth) and a chase between the wells were made on the PMMA plate. Each well had a through-hole in the bottom and Ag/AgCl electrodes were set into this hole (Fig. S1a†). A polymeric film made of parylene C (polychloro-*p*-xylylene) with a thickness of 5 μm was patterned with a single pore (100 μm diameter) using a conventional photolithography method and then fixed between the PMMA films (0.2 mm thick) using an adhesive bond (Super X, Cemedine Co., Ltd, Tokyo, Japan). The films, including the parylene film, were inserted into the chase to separate the wells.

Lipid bilayer preparation and reconstitution of αHL

Lipid bilayers were prepared using the microdevice (Fig. S1a†). Lipid bilayers can be simultaneously formed in this device by the droplet contact method[46–48] (Fig. S1b†). In this method, the two lipid monolayers contact each other and form lipid bilayers on a parylene C film that separates the two chambers. Lipid bilayers were formed as follows: the wells of the device were filled with *n*-decane (1.5 μL) containing DPhPC (10 mg mL^{-1}). The buffer solution (4.7 μL) with αHL (final concentration 30 nM) and DNA were poured into one chamber which was connected to the ground terminal. The buffer solution (4.7 μL) was also poured into another chamber which was connected to the recoding terminal. In this study, the buffer solution (0.5 M KCl and 5 mM MOPS (pH 7.0)) was used for each droplet. Within a few minutes of adding the solutions, a lipid bilayer formed and αHL formed a nanopore by reconstitution in the lipid bilayer. When the lipid bilayers ruptured during this process, they were reassembled by tracing with a hydrophobic stick at the droplet interface.

Channel current measurement and data analysis

The channel current was recorded using an Axopatch 200B amplifier (Molecular Devices, USA), filtered using a low-pass Bessel filter at 10 kHz at a sampling rate of 50 kHz. A constant voltage of +120 mV was applied from the recoding side, and the ground side was grounded. The recorded data from Axopatch 200B were acquired with the Clampex 9.0 software (Molecular Devices, USA) using a Digidata 1440A analog-to-digital converter (Molecular Devices, USA). Data were analyzed using Clampfit 10.6 (Molecular Devices, USA), Excel (Microsoft, Washington, USA), Python (Python Software Foundation, Delaware, USA), and Origin pro 8.5J (Light Stone, Tokyo, Japan). The analyzed data were obtained from over three different αHL nanopores. DNA translocation was detected when >80% of open αHL channel currents were inhibited.[40,49] The unzipping time was filtered between 99 ms and 25000 ms. The bootstrap method is based on the resampling of the original random sample drawn from a population with an unknown distribution. We used the exact bootstrap method, which availed the entire space of resamples. In the exact bootstrap method, accuracy verification is possible when the sample number is over 30.[50] In this study, our bootstrap procedure took 300 samples randomly from the primary common translocation data with 65 536 replacements and calculated the mean for these samples. The bootstrapped data were used to produce the histograms of the unzipping time. The error bar in Fig. 3h is the width of one standard deviation after bootstrapping.

Graduated PCR

For graduated PCR, the PCR mixture in a total volume of 20 μL contained: 10 μL of KOD SYBR® qPCR Mix, the extracted DNA, 0.2 μM of the forward primer, and 0.2 μM each node strand as the reverse primer. PCR thermal cycles were carried out using a thermal cycler (astec) to amplify the extracted DNA by the following process: preheat at 98 °C for 1 min; 45 cycles of 98 °C for 10 s, 58 °C for 10 s, 68 °C for 30 s. The result of the graduated PCR was confirmed by 10% non-denaturing polyacrylamide gel electrophoresis (containing 19/1 acrylamide/bis (w/w)) in 1× TBE buffer at a constant power of 2.3 W for 90 min at room temperature. The gel was prepared in our laboratory. After electrophoresis, the gel was stained with diluted SYBR Green II (Takara Bio Inc., Japan) solution for 30 min and visualized under blue LED irradiation; images were obtained using an LED transilluminator (Bio Craft Co., Ltd, Tokyo, Japan).

Author contributions

S. T. and R. K. conceived the original idea and S. T. performed the experiments. R. K. supervised the research. S. T. and R. K. wrote the entire manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was partly supported by a Grant-in-Aid for Scientific Research (KAKENHI) (Grant No. 19H00901, 17K19138, and 15H00803) from the Ministry of Education, Culture, Sport, Science and Technology (MEXT), Japan. The authors thank A. Cooney for language editing.

References

1 L. M. Adleman, *Science*, 1994, **266**, 1021–1024.
2 S. Y. Hsieh, C. W. Huang and H. H. Chou, *Appl. Math. Comput.*, 2008, **203**, 502–512.
3 R. J. Lipton, *Science*, 1995, **268**, 542–545.
4 D. Faulhammer, A. R. Cukras, R. J. Lipton and L. F. Landweber, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 1385–1389.

5 Y. Benenson, T. Paz-Elizur, R. Adar, E. Keinan, Z. Livneh and E. Shapiro, *Nature*, 2001, **414**, 430–434.
6 R. S. Braich, N. Chelyapov, C. Johnson, P. W. K. Rothemund and L. Adleman, *Science*, 2002, **296**, 499–502.
7 J. Y. Lee, S. Y. Shin, T. H. Park and B. T. Zhang, *BioSystems*, 2004, **78**, 39–47.
8 F. Tanaka, A. Kameda, M. Yamamoto and A. Ohuchi, *Nucleic Acids Res.*, 2005, **33**, 903–911.
9 F. S. Xiong, D. Spetzler and W. D. Frasch, *Integr. Biol.*, 2009, **1**, 275–280.
10 M. Yamamoto, A. Kameda, N. Matsuura, T. Shiba, Y. Kawazoe and A. Ohuchi, *New Gener. Comput.*, 2002, **20**, 251–261.
11 Q. H. Liu, L. M. Wang, A. G. Frutos, A. E. Condon, R. M. Corn and L. M. Smith, *Nature*, 2000, **403**, 175–179.
12 F. S. Xiong and W. D. Frasch, *Nat. Comput.*, 2011, **10**, 947–959.
13 Z. Ibrahim, J. A. Rose, A. Suyama and M. Khalid, *Nat. Comput.*, 2008, **7**, 277–286.
14 L. Q. Gu, O. Braha, S. Conlan, S. Cheley and H. Bayley, *Nature*, 1999, **398**, 686–690.
15 H. Bayley, B. Cronin, A. Heron, M. A. Holden, W. L. Hwang, R. Syeda, J. Thompson and M. Wallace, *Mol. BioSyst.*, 2008, **4**, 1191–1208.
16 J. E. Reiner, A. Balijepalli, J. W. F. Robertson, J. Campbell, J. Suehle and J. J. Kasianowicz, *Chem. Rev.*, 2012, **112**, 6431–6451.
17 D. H. Stoloff and M. Wanunu, *Curr. Opin. Biotechnol.*, 2013, **24**, 699–704.
18 C. Cao and Y. T. Long, *Acc. Chem. Res.*, 2018, **51**, 331–341.
19 T. L. Ding, J. Yang, V. Pan, N. Zhao, Z. H. Lu, Y. G. Ke and C. Zhang, *Nucleic Acids Res.*, 2020, **48**, 2791–2806.
20 Y. L. Ying and Y. T. Long, *J. Am. Chem. Soc.*, 2019, **141**, 15720–15729.
21 S. M. Lu, Y. Y. Peng, Y. L. Ying and Y. T. Long, *Anal. Chem.*, 2020, **92**, 5621–5644.
22 L. Z. Song, M. R. Hobaugh, C. Shustak, S. Cheley, H. Bayley and J. E. Gouaux, *Science*, 1996, **274**, 1859–1866.
23 J. J. Kasianowicz, E. Brandin, D. Branton and D. W. Deamer, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 13770–13773.
24 A. Meller, L. Nivon, E. Brandin, J. Golovchenko and D. Branton, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 1079–1084.
25 A. Meller, L. Nivon and D. Branton, *Phys. Rev. Lett.*, 2001, **86**, 3435–3438.
26 D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. H. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. S. S. Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Pershin, J. M. Ramsey, R. Riehn, G. V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin and J. A. Schloss, *Nat. Biotechnol.*, 2008, **26**, 1146–1153.
27 D. Stoddart, A. J. Heron, J. Klingelhoefer, E. Mikhailova, G. Maglia and H. Bayley, *Nano Lett.*, 2010, **10**, 3633–3637.
28 M. N. Stojanovic and D. Stefanovic, *Nat. Biotechnol.*, 2003, **21**, 1069–1074.
29 G. Seelig, D. Soloveichik, D. Y. Zhang and E. Winfree, *Science*, 2006, **314**, 1585–1588.
30 L. Qian, E. Winfree and J. Bruck, *Nature*, 2011, **475**, 368–372.
31 D. Y. Zhang and G. Seelig, *Nat. Chem.*, 2011, **3**, 103–113.
32 L. Qian and E. Winfree, *Science*, 2011, **332**, 1196–1201.
33 H. Yasuga, R. Kawano, M. Takinoue, Y. Tsuji, T. Osaki, K. Kamiya, N. Miki and S. Takeuchi, *PLoS One*, 2016, **11**, e0149667.
34 H. Yasuga, K. Inoue, R. Kawano, M. Takinoue, T. Osaki, K. Kamiya, N. Miki and S. Takeuchi, *PLoS One*, 2017, **12**, e0180876.
35 M. Hiratani, M. Ohara and R. Kawano, *Anal. Chem.*, 2017, **89**, 2312–2317.
36 M. Ohara, M. Takinoue and R. Kawano, *ACS Synth. Biol.*, 2017, **6**, 1427–1432.
37 R. Kawano, *Biotechnol. J.*, 2018, **13**, 1800091.
38 M. Hiratani and R. Kawano, *Anal. Chem.*, 2018, **90**, 8531–8537.
39 P. Liu and R. Kawano, *Small Methods*, 2020, **4**, 2000101.
40 R. T. Perera, A. M. Fleming, A. M. Peterson, J. M. Heemstra, C. J. Burrows and H. S. White, *Biophys. J.*, 2016, **110**, 306–314.
41 A. F. Sauer-Budge, J. A. Nyamwanda, D. K. Lubensky and D. Branton, *Phys. Rev. Lett.*, 2003, **90**, 238101.
42 M. Ohara, Y. Sekiya and R. Kawano, *Electrochemistry*, 2016, **84**, 338–341.
43 H. L. Zhang, M. Hiratani, K. Nagaoka and R. Kawano, *Nanoscale*, 2017, **9**, 16124–16127.
44 K. K. Chen, J. B. Zhu, F. Boskovic and U. F. Keyser, *Nano Lett.*, 2020, **20**, 3754–3760.
45 M. Hagiya, A. Konagaya, S. Kobayashi, H. Saito and S. Murata, *Acc. Chem. Res.*, 2014, **47**, 1681–1690.
46 R. Kawano, *ChemPhysChem*, 2018, **19**, 359–366.
47 Y. Tsuji, R. Kawano, T. Osaki, K. Kamiya, N. Miki and S. Takeuchi, *Anal. Chem.*, 2013, **85**, 10913–10919.
48 R. Kawano, Y. Tsuji, K. Sato, T. Osaki, K. Kamiya, M. Hirano, T. Ide, N. Miki and S. Takeuchi, *Sci. Rep.*, 2013, **3**, 1995.
49 T. Z. Butler, J. H. Gundlach and M. Troll, *Biophys. J.*, 2007, **93**, 3229–3240.
50 J. Kisielinska, *Comput. Stat.*, 2013, **28**, 1061–1077.
51 Y. Yang, R. Zhang and C. H. Fan, *Trends Chem.*, 2020, **2**, 137–147.
52 H. Ramezani and H. Dietz, *Nat. Rev. Genet.*, 2020, **21**, 5–26.
53 N. A. W. Bell and U. F. Keyser, *FEBS Lett.*, 2014, **588**, 3564–3570.