DNA Structural Barcode Copying and Random Access

Filip Bošković, Alexander Ohmann, Ulrich F. Keyser,* and Kaikai Chen*

1. Introduction

Deoxyribonucleic acid (DNA) nanotechnology offers structural control with sub-nanometer precision to build designed nanostructures by DNA self-assembly relying on Watson–Crick base pairing.[2,4] This striking advantage was exploited in numerous ways to create different functional DNA nanostructures.[1,3] One useful aspect that DNA nanotechnology offers is to precisely position DNA nanostructures along a DNA strand in a programmable arrangement.[2,4] By annealing a several kilobase-long, single-stranded (ssDNA) “scaffold” strand with short (38-48 nt) complementary oligonucleotides containing DNA nanostructures, uniquely addressable DNA structural barcodes can be created.[2] It is comparable to the DNA sequence barcode which has been used to specifically label DNA fragments for DNA sequencing and genomic library preparation.[5,6] However, with the advance in DNA nanotechnology, DNA structural barcodes emerged as a potential alternative by encoding information in the structures rather than in the sequences. They are more robust in carrying embedded information compared with the DNA sequence barcodes, as multiple single-nucleotide changes could cause misinterpretation of DNA sequence barcode,[7] whereas the barcode readout of the 3D structures will not be affected.

Furthermore, it is easier and more accurate to read the codes on the DNA structures using nanopore-based methods because they can be designed with a more significant difference in size compared with the small difference in single nucleotides. Some of the versatile applications of DNA structural barcodes are data storage in DNA nanostructures[8] and multiplexed sensing of single proteins both in vitro[2] and in vivo.[9] DNA barcodes have been shown as promising tools for multiplex intracellular protein sensing.[9] Combined with fluorescent labeling, they have additionally shown tremendous applications in the field of DNA-based point accumulation for imaging in nanoscale topography (DNA-PAINT) to create a fluorescent barcode.[10,11] DNA structural barcodes have also a great potential to be used in whole-genome mapping strategies, such as single-molecule optical mapping.[12]

To read out the information stored in DNA structural barcodes, nanopore-based single-molecule sensing has proven to be a powerful tool for numerous applications.[2,8,13] The principle of nanopore sensing is based on passing a single molecule through a small confinement by applying a voltage across it. The translocation of the charged molecule driven by the electric field creates a current blockade which is determined by the molecule’s shape, volume, and charge allowing us to read out structural information in a label-free manner. Recently, we have pushed this even further using DNA nanostructures for an innovative data-storage approach as an alternative to data storage in DNA sequence[14] including the ability to rewrite and securely store data in so-called DNA hard drives.[15] DNA structural barcodes are not easy to be copied because of their complexity[14,15] compared with a DNA sequence. Previously, a few attempts to replicate DNA nanostructures were demonstrated either via plasmid replication in bacteria[16,17] or by self-replicating DNA materials.[18,19] From these approaches, DNA nanostructure de novo copying relies on in vivo replication of plasmid-encoded sequence information.[16,17] Although DNA structural barcodes may be copied via transformation of bacteria

F. Bošković, Dr. A. Ohmann, Prof. U. F. Keyser, Dr. K. Chen
Cavendish Laboratory
JJ Thomson Avenue, Cambridge CB3 0HE, UK
E-mail: ufk20@cam.ac.uk; kc494@cam.ac.uk

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/sstr.202000144.

© 2021 The Authors. Small Structures published by Wiley-VCH GmbH

DOI: 10.1002/sstr.202000144

www.small-structures.com
with the self-replicating unit, an approach based on polymerase chain reaction (PCR) amplification is more valuable as it is a standardized laboratory method. However, a key challenge is the asymmetry of the DNA structural barcode design. Although the scaffold strand is linear, its complement contains individual oligonucleotides where some of them include 3D DNA protrusions. A DNA structural barcode composed of multiple, closely spaced DNA nanostructures contains repetitive DNA.

In addition, accessing a desired DNA nanostructure barcode in a complex mixture of similar barcodes is still a huge challenge because of their similar design and usage of the same DNA scaffold. Random access is of paramount significance for targeting and enrichment of the DNA structural barcode of interest, for example, for only reading desired data files but not the whole archive of stored data. As single-molecule methods typically require long measurement time or multiple measurements, significant amounts of library material could be wasted to detect one file. Hence, random access is critical for more realistic applications of DNA structural barcodes carrying digital information.

To overcome those emerging challenges, we successfully developed barcode copying and random access of single barcodes from the library of DNA structural barcodes in this study. First, we designed and assembled a library of DNA structural barcodes where a structural bit is composed of multiple dumbbell-shaped, double-hairpin DNA protrusions (see Figure 1A and Figure S1, S2, Supporting Information for details). The design of the barcodes was verified with nanopore sensing and the expected barcode lengths were confirmed with agarose gel electrophoresis. Subsequently, we successfully copied a barcode with 6 bits with exponential PCR amplification. Finally, we designed a library of barcodes with an end-specific primer that we used to access the desired barcode by barcode-specific linear PCR amplification and verified the output with nanopore sensing. Our study introduces an easy and affordable approach toward fabricating data storage archives and barcode libraries for various biomolecule-sensing purposes using DNA structural barcodes.

2. Results
2.1. DNA Structural Barcode Design, Assembly, and Readout

To create DNA structural barcodes, we first prepared ssDNA scaffold (7228 nt) by linearizing circular M13mp18 DNA and subsequently annealing this linear scaffold (Figure 1A left) with short, 5’ phosphorylated oligonucleotides (oligonucleotide sequences are listed in Table S1, Supporting Information). The particular barcode was then assembled by designing selected complementary oligonucleotides with a double-hairpin structure (Figure 1A middle) termed “DNA dumbbell” (sequences listed in Table S2–S5, Supporting Information). This DNA nanostructure is formed by two hairpins each consisting of 5 bp and 4 dT loop as shown in Figure S2, Supporting Information. Each single bit is constructed by closely spacing eight DNA dumbbell nanostructures together along the long DNA scaffold (Figure S1, Supporting Information) as they are big enough to cause a sufficient signal measured by the ~14 nm diameter nanopores used in this study. As characterized in a previous study, fewer dumbbells, for example, 5 dumbbells were also able to cause a signal using ~14 nm nanopores, and even smaller structures were observable with smaller nanopores. Here, we chose 8 as an example and this number and the nanopore size can be tuned to strike a balance between the signal strength and the number of bits required for specific applications. After annealing with the short oligonucleotides, we obtained a double-stranded DNA barcode with multiple DNA nicks (indicated by vertical black arrows in Figure 1A middle). We copied the structure using PCR by linking the separate oligonucleotides together by ligating the DNA nicks using T4 DNA ligase. The ligation forms a long continuous dsDNA strand with designed bits protruding from it (Figure 1A right).

The correct assembly of DNA structural barcodes was verified by reading out the barcode using nanopore sensing (Figure 1B). The negatively charged DNA structural barcodes

Figure 1. DNA structural barcode assembly and nanopore readout. A) DNA barcodes are assembled by hybridizing ssDNA scaffolds (7228 nt) with short complementary oligonucleotides (38–48 nt). Specific oligonucleotides are designed to form double DNA hairpins (DNA dumbbells). Each structural bit of the DNA barcode consists of eight, closely spaced DNA dumbbells. DNA nicks (black arrows), resulting from separate oligos hybridized to the scaffold, are ligated using a T4 DNA ligase to form the final double-stranded DNA structural barcode. B) A negatively charged DNA structural barcode passes through the nanopore creating a barcode-specific nanopore signal as the readout. The structural bit is detected as a downward peak in the current signal assigned as “1”.

© 2021 The Authors. Small Structures published by Wiley-VCH GmbH
were electrophoretically driven through the nanopore by the applied electric field thereby causing a measurable, barcode-specific modulation of the ionic current. The nanopore sensing method can detect protrusions on the DNA as an additional, distinct downward peak in current signal as shown in Figure 1B (right). The position of the bit in the design corresponds to the peak position in the ionic current recording due to a relatively constant DNA translocation velocity. The presence of a structural bit, indicated by a downward peak, is assigned as “1”, and the absence of a peak is assigned as a structural bit “0”.

2.2. Nanopore Microscope Verification of a Designed Structural Barcode Library

Before copying, we first verified each barcode design individually using nanopore sensing. For this, we collected datasets for each barcode design, isolated single barcode translocations along the current trace, and used only unfolded barcode translocations for further analysis (for detailed analysis workflow see Section 3, Supporting Information). We prepared four barcode designs: one barcode with 6 bits and another three each with 3 bits (Figure 2). Figure 2A shows the 6-bit DNA barcode design together with an example of the corresponding nanopore signal. Each bit is highlighted in red and corresponds to a downward peak in the nanopore recording. For \(\approx 93\%\) of the events, 6 bits corresponding to our designed barcode could be determined (Figure 1A), demonstrating the compelling accuracy of our approach.

For the 3-bit systems, we designed three DNA structural barcodes encoding for “111”, “011”, and “001”, by having three, two, or one structural unit, respectively. Illustrations of barcodes “111”, “011”, and “001” are shown in Figure 2 respectively, with representative events and the statistical analysis of the assigned readout to a certain barcode. Additional examples are shown in Figure S4, Supporting Information, with corresponding event statistics in Figure S8, Supporting Information. All three designs demonstrated a readout with >90% correct, further establishing the robustness of our method. An additional feature of 3-bit barcodes is the non-complementary ends (Table S6, Supporting Information) serving as a template for barcode-specific copying and access which will be discussed later in this study.

2.3. Copying and Quantification of DNA Structural Barcodes

We demonstrate the copying capability by amplifying the barcode “111111” with 6 bits introduced in Figure 2A using LongAmp Taq DNA Polymerase (NEB). Using a thermocycler for the PCR reaction, the double-stranded DNA barcode was denatured by heat to obtain two ssDNA strands. These were subsequently annealed with forward and reverse primers (for sequences, see Table S2, Supporting Information) followed by elongation step of the primer strands. Before copying, the majority of events contained six structural units and the barcode consists of two strands B (containing DNA dumbbells) and S (DNA scaffold), as verified in Figure 2A. In the first amplification cycle, strands B and S were copied to obtain the dsDNA constructs BB and SS (B' and S' are the complementary strands of B and S, respectively) using the forward and the reverse primers. After the additional 29 cycles of exponential PCR copying of initial asymmetric DNA strands, four different strand combinations could exist. BS and B’S’ should have identical ionic current nanopore traces with the correct reading of “111111” while the SS’ combination consists.

![Figure 2. DNA structural barcode characterization. We designed and tested four different barcode signatures represented by a varying number of bits (red). A) Barcode “111111” contains six equally spaced bits. Exemplary nanopore readout showing six peaks in the current signal is depicted below. 93% of counted events were correctly assigned with 6 bits. B-D) Three different barcodes containing three (B), two (C), and one (D) bit were verified with nanopore barcode readout yielding 93%, 90%, and 97% accuracy, respectively. Exemplary events are depicted below each design. Non-complementary terminal sequences offer barcode-specific amplification using strand-specific primers (discussed in Figure 4).](image-url)
of complementary DNA strands without the bits (Figure 3A). The last combination is BB' that might form dsDNA constructs with or without protrusions due to the full complement of protrusions.

DNA structural barcodes were verified by nanopore measurements before and after copying confirming that the six structural bits were still preserved after PCR copying (Figure 3B). Importantly, we were able to amplify the initial 5 ng of original barcodes by close to around 1000-fold, yielding 4.8 μg of copied DNA barcode structures. Statistical nanopore single-molecule readouts confirmed that the majority (>60%) of counted events were still correctly assigned. Additional examples of the correct barcode readout after amplification are shown in Figure S6, Supporting Information. As it can be seen from the example events (the first 20 events, Figure S9, Supporting Information), barcode readout was correctly achieved after PCR copying.

To support the single-molecule results of the nanopore experiments, we additionally performed agarose gel electrophoresis as a bulk analysis method to verify the DNA barcode lengths before and after copying. By comparing the observed gel bands to a DNA ladder, we could confirm that the DNA barcodes before and after amplification agree with the expected length of close to 8 kb (Figure 3C), thus furthermore demonstrating the successful copying process.

2.4. Random Access of the Desired Target Barcode from a Barcode Library

As introduced earlier, the 3-bit barcode structures were designed to contain a non-complementary, unique end sequence which allows us to use a barcode-specific primer to access a single barcode from the library of mixed barcodes (Figure 4A). We demonstrate the access of the desired target barcode from a library by amplying one barcode from an equimolar mixture of three 3-bit barcodes. For this, barcodes “111”, “011”, and “001” were mixed in equimolar concentrations. To perform the linear PCR amplification, we added barcode “111”-specific primer strand and the “M13mp18 forward primer”. After performing 60 amplification cycles, the output of this reaction should be mainly composed of the amplified barcode “111”.

We validated this by nanopore measurements before and after copying of barcode “111” from the described barcode library (Figure 4B; see Figure S5, S7, Supporting Information, for example events). Before amplification, all barcodes were read out with an equal probability (around 30%), as expected from an equimolar barcode mixture. However, after amplification, the vast majority (>88%) of barcodes were assigned to the desired amplified barcode “111”. Three representative events before and after amplification are shown in Figure 4B. All three different barcodes were detected before the amplification, but after...
Selective copying, barcode “111” was in excess as quantitatively estimated from a NanoDrop spectrophotometer (Figure 4B). Furthermore, we used agarose gel electrophoresis again to confirm that the barcodes display the expected band at approximately 8 kb before and after selective amplification (Figure 4C). The little difference between the positions of the “copies” and those of the original ones might be caused by the stacking interactions between DNA molecules or the difference in the buffer condition.

3. Discussion

In this study, we demonstrated that our DNA structural barcode designs can be successfully copied which we verified by single-molecule nanopore analysis and bulk agarose gel electrophoresis analysis. Second, we demonstrated that by using barcode-specific primers, we can randomly access a specific target barcode from a library of constructs.

We successfully copied the 6-bit DNA structural barcode using exponential PCR that should create an equimolar mixture of four DNA strands that can hybridize into multiple copies of the original barcode “111111”. Still, the complexity of the PCR reaction can lead to a different number of copied DNA dumbbells (= structural units) due to the repetitive nature of the sequence forming DNA protrusions.[22] However, we show that our approach is robust and that the barcode readout is correct regardless of imperfections of PCR amplification.

As a proof-of-concept, we used DNA dumbbells (double DNA hairpins). However, a DNA protrusion can be replaced with any nanopore signal-inducing secondary structures such as DNA hairpins,[8] G-quadruplexes,[23] or a non-complementary DNA loop that contains a binding site for additional DNA nanostructures[24] or biotinylated strands.[13] In addition, the

Figure 4. Random access to a specific DNA structural barcode by selective copying. A) DNA structural barcode library represents the equimolar mixture of barcodes “111”, “011”, and “001”. Barcode-specific non-complementary ends are highlighted in different colors. Random access is achieved by selective PCR copying using a barcode-specific pair of primers. As an example, barcode “111” is amplified for 60 cycles, with DNA amount indicated in nanograms. B) DNA structural barcode library including exemplary current traces before and after copying highlighting the increased nanopore readout of barcode “111” after its selective amplification from an equimolar barcode mixture. C) Agarose gel electrophoresis verifies the size of each barcode design before copying as well as in the purified and diluted mixture after PCR. 1 kb ladder indicates that barcodes are close to 8 kb.
binding site can be in the loop region of the DNA stem-loop structure and both structures could be copied and accessed in a similar way as DNA dumbbells and serve as additional ways to create DNA structural barcodes. As shown previously, one can rewrite such barcode that further can be used as a unique on-demand programmable platform for barcoding.[2,13]

Self-assembled structural DNA barcodes are one of the most promising choices to create durable barcodes for nanopore sensing as they have several advantages. The barcode readout in nanopores should not be affected if several DNA dumbbells are missing.[2] DNA nanostructures are intrinsically unaffected by a few nucleotide substitutions in the sequence of the barcode because this would not cause a significant difference in the barcode structure. Using DNA self-assembly or strand displacement reactions one can easily change the position and number of DNA structures in a barcode, as we previously demonstrated.[2,13]

Random access in DNA data libraries is of particular interest for DNA data storage.[25] Our work will pave the way for data storage in DNA nanostructures—a new approach shown recently as an alternative to DNA sequence-based data storage.[13,26–28] The structural DNA barcode method has both advantages in data reading (multiplexed nanopore sensing) and writing (parallel writing). For the writing, enzyme-free mixing of the oligonucleotides is easier than synthesizing oligos.[13] The high-throughput production of DNA structural barcodes could be achieved by either using digital microfluidics[29] or nanoliter piezo dispensers,[30] with the potential of operating a large number of droplets in parallel with physical separation.

The same system can be used to access a stored data file of interest by copying it. Large-scale production of DNA structural barcodes would be needed for structure-based DNA tags for tracking applications.[31] We have demonstrated that nanopore sensing might be an alternative readout method for DNA tags in comparison to nanopore sequencing.[32] We created multiple copies of a specific data file written in a DNA structural annotation that can expand the potential of DNA nanostructures and nanopores as a feasible way to archive and access data.

Our study will propel the field toward the sustainable, larger-scale application of DNA structural barcodes for multiplexed single-molecule sensing of biomolecules as has been introduced previously.[2] Production of larger quantities of such barcodes has been one of the limiting factors for industrial-scale applications of self-assembled DNA systems.[33,34]

4. Experimental Section

DNA Structural Barcode Assembly: We fabricated the barcodes by annealing a long linear ssDNA scaffold (Guild BioSciences) with short complementary oligonucleotides (Integrated DNA Technologies, Inc.) of which some had a fully complementary sequence (Table S1, Supporting Information). A select number of these oligos additionally contained DNA protrusions at specific positions (Figure 1A, Table S2–S5, Supporting Information). The scaffold linearization protocol was previously described.[2] The oligonucleotide set for assembling a specific barcode design was prepared by mixing the required oligonucleotides (200 nM final concentration per oligonucleotide) and phosphorylating the 5’ ends of the oligonucleotides using the T4 polynucleotide kinase (PNK) kit (New England Biolabs [NEB]). For phosphorylation of oligonucleotides, 50 µL of 1000 pico moles of oligonucleotides were mixed with 10 µL 10× ligase buffer, 10 µL 10 mM ATP, 7 µL T4 PNK, and 23 µL of nuclease-free water. The mix was incubated overnight at 37 °C and heat-inactivated at 65 °C for 20 min.

The assembly mix for DNA structural barcodes was composed of: 8 µL linearized M13mp18 DNA scaffold (100 nM); 40 µL phosphorylated oligonucleotide mix (100 nM); 4 µL 100 mM MgCl2; 1.2 µL 100 mM Tris-HCl (pH 8.0); 10 mM EDTA; 6.8 µL Mill-Q water. The mixture was then heated to 70 °C followed by a linear cooling ramp to 25 °C over 50 min. Immediately after the assembly, annealed barcodes were purified from excess oligonucleotides via spin filtration using 100 kDa Amicon filters by mixing the annealed mixture and adding washing buffer (10 mM Tris-HCl pH 8.0, 0.5 mM MgCl2) up to 500 µL and centrifugated at 9000 × g for 10 min. This step was repeated twice. The purified DNA structural barcode was retrieved by reversing the filter and centrifugation for 2 min at 10000 × g.

Finally, DNA nicks left between hybridized oligonucleotides were ligated using the T4 DNA ligation kit (NEB) by mixing 100 ng of purified barcodes with 2 µL of 10× ligase buffer, 2 µL of T4 DNA ligase, and filled up to 20 µL with nuclease-free water. The ligation reaction was incubated overnight at 16 °C, followed by inactivation at 65 °C for 10 min. The product of the ligation reaction was purified using the Monarch PCR and DNA Clean up Kit (5 µg), according to the manufacturer’s instructions.

Copying of DNA Structural Barcodes with Exponential Amplification: To demonstrate the exponential barcode PCR amplification, we used a barcode design with 6 bits (“111111”). Exponential PCR amplification was performed using the LongAmp Taq PCR kit (NEB; for primer sequences, see Table S7, Supporting Information) at an initial DNA amount of ~5 ng per reaction as estimated using a NanoDrop spectrophotometer. Samples were initially denatured at 94 °C for 4 min and subsequently amplified in 30 cycles of 94 °C (30 s), 54 °C (30 s), and 72 °C (7.5 min; ~50 s kb−1). The elongation time was adapted to the length of the strand containing DNA dumbbells according to a rate of ~1 kb per second. After the amplification cycles, a final extension of 72 °C for 10 min was performed, followed by storage at 4 °C. Samples were then purified using the Monarch PCR and DNA Cleanup Kit (5 µg).

Random Access to Target DNA Structural Barcode with Linear Amplification: To demonstrate the linear barcode PCR amplification, we used 3-bit structural barcodes with unique terminal end overlaps. These terminal ends (for terminal end sequence, see Table S6, Supporting Information) were barcode-specific and were used as primer-binding sites. Linear PCR amplification was performed using the LongAmp Taq PCR kit (NEB; for primer sequences, see Table S7, Supporting Information) at an initial DNA amount of ~10 ng per reaction as estimated using a NanoDrop spectrophotometer. Samples were initially denatured at 94 °C for 4 min and subsequently amplified in 60 cycles of 94 °C (30 s), 54 °C (30 s), and 72 °C (7.5 min; ~50 s kb−1). The elongation time was adapted to the length of the strand containing DNA dumbbells according to a rate of ~1 kb per second. After the amplification cycles, a final extension of 72 °C for 10 min was performed, followed by storage at 4 °C. Samples were then purified using the Monarch PCR and DNA Cleanup Kit (5 µg).

Agarose Gel Electrophoresis: Lengths of the barcode constructs were verified using 0.8% (w/v) agarose (Sigma-Aldrich, BioReagent) gel electrophoresis in a 0.5 × Tris–borate–EDTA (TBE, pH 8.0) buffer. 150 ng sample was loaded using a 6 × SDS-free, purple loading dye (NEB). The gel was run for 90 min at 4 °C at 60 V, post-stained in 3 × GelRed (Biotium), and imaged using a GelDoc-it UV imaging system. A 1 kb ladder (NEB) was used in all experiments. The grayscale was inverted in the ImageJ software and the background was subtracted using the integrated rolling-ball method at a radius of 150 pixels.

Nanopore Measurement and Data Analysis: To detect the designed barcodes and their bits, we used 14 ± 3 nm glass nanopores fabricated by pulling quartz capillaries with filaments (0.5 mm outer diameter and 0.2 inner diameter, Sutter Instrument, USA) using a laser-assisted puller (P-2000, Sutter Instruments).[21]

The details of nanopore measurement, nanopore setup, and data analysis were used as described previously (Figure S3, Supporting Information).[21]
Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors thank Pablo A. Vargas Rosales for critical reading of the manuscript. Funding: U.F.K. and K.C. acknowledge funding from a European Research Council (ERC) consolidator grant (DesignerPores No. 647144). F.B. acknowledges funding from George and Lilian Schiff Foundation Studentship, the Winton Programme for the Physics of Sustainability Ph.D. Scholarship, and St John’s College Benefactors’ Scholarship. A.O. acknowledges funding from the Cambridge Trust Vice Chancellor’s Award.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

deoxyribonucleic acid data storage, deoxyribonucleic acid nanostructures, deoxyribonucleic acid nanotechnology, nanopores, single-molecule

Received: January 4, 2021
Revised: January 29, 2021
Published online: February 22, 2021