Synthetic Biology, 2025, **10(1)**, ysaf008
DOI: https://doi.org/10.1093/synbio/ysaf008
Advance Access Publication 7 April 2025
Short Communication

Engineering a custom-sized DNA scaffold for more efficient DNA origami-based nucleic acid data storage

Sarah E. Kobernat[1,2,3], Maryna Lazouskaya[4], Benjamin C. Balzer[2,3], Amanda Wolf[2,3], Golam M. Mortuza[5], George D. Dickinson[3], Tim Andersen[5], William L. Hughes[3,6], Luca Piantanida[3,6], Eric J. Hayden[1,2,3,*]

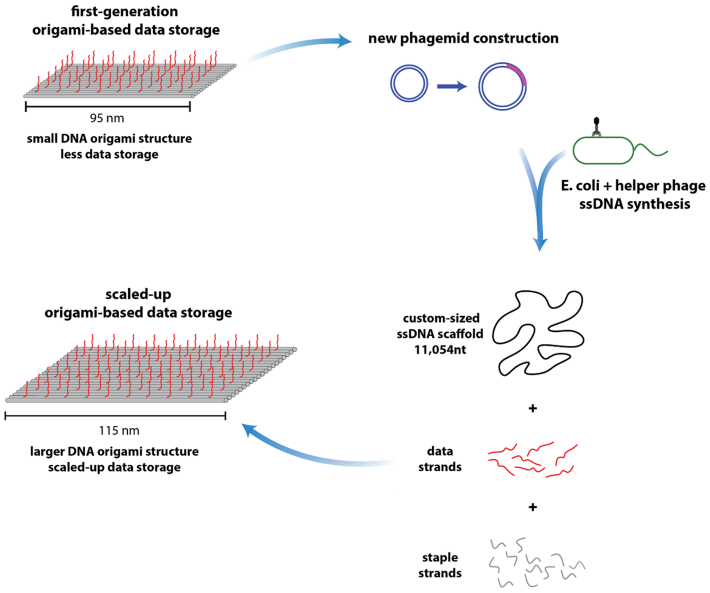
[1]Biomolecular Sciences Graduate Programs, Boise State University, Boise, ID 83725, United States
[2]Department of Biological Sciences, Boise State University, Boise, ID 83725, United States
[3]Micron School of Materials Science & Engineering, Boise State University, Boise, ID 83725, United States
[4]Department of Electrical and Computer Engineering, Boise State University, Boise, ID 83725, United States
[5]Department of Computer Science, Boise State University, Boise, ID 83725, United States
[6]Faculty of Applied Science, School of Engineering, University of British Columbia, Kelowna, BC V1V 1V7, Canada
*Corresponding author. Biomolecular Sciences Graduate Programs, Boise State University, Boise, ID 83725, United States. E-mail: erichayden@boisestate.edu.

Abstract

DNA has emerged as a promising material to address growing data storage demands. We recently demonstrated a structure-based DNA data storage approach where DNA probes are spatially oriented on the surface of DNA origami and decoded using DNA-PAINT. In this approach, larger origami structures could improve the efficiency of reading and writing data. However, larger origami require long single-stranded DNA scaffolds that are not commonly available. Here, we report the engineering of a novel longer DNA scaffold designed to produce a larger rectangle origami needed to expand the origami-based digital nucleic acid memory (dNAM) approach. We confirmed that this scaffold self-assembled into the correct origami platform and correctly positioned DNA data strands using atomic force microscopy and DNA-PAINT super-resolution microscopy. This larger structure enables a 67% increase in the number of data points per origami and will support efforts to efficiently scale up origami-based dNAM.

Keywords: DNA origami; DNA data storage; nucleic acid memory; ssDNA synthesis; DNA nanotechnology

Graphical Abstract



Submitted: 3 April 2025; **Received (in revised form)**: 14 January 2025; **Accepted**: 6 April 2025

1. Introduction

The growing demand for data storage is quickly outpacing our current technologies. Due to increased rates of data generation and a trend of storing data for longer periods of time, there has been a projected compound annual growth rate of up to 50% from 2020 to 2030 [1]. The total amount of stored data globally is projected to reach ~180 zettabytes by 2025 and is predicted to outstrip the world's supply of silicon by 2040 [2, 3]. In addition to storage capacity, there is growing demand to reduce the total cost and energy consumption associated with storing data, including the energy required for manufacturing and maintaining hardware, and for writing, transferring, and retrieving data [4]. New materials and technologies are needed to sustainably meet the world's data storage demands.

DNA is a promising data storage material because of its inherent data density, chemical stability, and low energy demands [5, 6]. The majority of DNA data storage approaches have been designed around DNA sequencing technologies. These approaches generally encode data in nucleotide sequences that are synthesized in a collection of DNA oligonucleotides that can be stored and later recovered by high-throughput sequencing platforms. While these approaches are promising, there are also challenges due to errors associated with DNA synthesis, amplification, and sequencing, and the approach is tied to advancements in DNA sequencing platforms [7]. An alternative approach to DNA data storage is to use the self-assembling properties of DNA to create nanostructures capable of encoding data. We recently reported an approach, termed digital nucleic acid memory (dNAM) [8] that uses a rectangular DNA origami nanostructure [9] to position DNA strands protruding from the surface. The presence or absence of a protruding strand represents the binary state of that particular site. Upon addition of a fluorescent DNA probe, the binary data matrix can be read using a super-resolution microscopy (SRM) approach called DNA-PAINT (Point Accumulation for Imaging in Nanoscale Topography) [10]. The fluorescent pattern of each DNA origami is encoded and decoded with a custom algorithm.

The quantity of data that can be stored in one dNAM origami structure, or 'node', is limited by the size of the origami surface. Larger DNA origami structures can contain more bits, use more complex encoding schemes, and potentially reduce origami synthesis and reading effort and time. However, the size of the DNA origami structures used in our dNAM approach is limited by the size of the DNA 'scaffold', which is the long single-stranded DNA (ssDNA) molecule used in hybridization [11, 12]. Similar to a lot of DNA origami research, our dNAM prototype used the commercially available 7249 nucleotide (nt) long single-stranded DNA scaffold isolated from the phage vector M13mp18. The length of this common scaffold limited the size of the origami rectangle to ~95 nm × 70 nm in size (Fig. 1a). Building larger origami structures to improve dNAM requires the design and synthesis of larger ssDNA scaffolds. In addition, larger scaffolds need to be tested for production in *Escherichia coli*-based systems and for accurate folding into origami capable of precisely orienting ssDNA data strands.

Here, we report the custom design, synthesis, production and characterization of a longer ssDNA scaffold engineered for dNAM. We first designed a larger origami rectangle that could add two additional rows and columns of data to each nanostructure, while maintaining the same distance between data strands (Fig. 1b). Based on this origami structure, we designed the ssDNA scaffold that was 11 054 nucleotides long and precisely fits our designed structure. To synthesize the scaffold, we designed and constructed a novel phagemid that could express the desired scaffold upon transformation into *E. coli* and infection with helper phage (Fig. 1c). The scaffold was designed to assemble an ~116 nm × 95 nm rectangular structure. Using the same 10-nm spacing of data strands, the new larger origami achieves a 67% increase in data points per origami relative to our prototype (Fig. 1).

2. Materials and methods

2.1 DNA origami design

An ~116 nm × 95 nm rectangular DNA origami structure was designed in caDNAno using the square lattice similar to previously described structures [13]. The structure is comprised of 32 helices each 352 bp in length. This structure requires a 11 054-nt scaffold with 352 staple strands ~30 nt long (for detailed sequences, see Supplementary File S1). Additional sequence 5′-TTGGGAGGA-3′ were added to select staples to create protruding strands at desired locations.

2.2 Cloning

Restriction enzyme cloning was used to add a 982 bp insert to pScaf-10080.1. The insert was PCR amplified from pScaf-3024.1 (Addgene plasmid #111404) with primers F-5′-ggtaccagtctcacctcaagaagctgtttgctgg-3′ and R-5-agactccttacccaacagctcgcagacggt-3′ that added KpnI and BglII restriction sites to the 5′ and 3′ ends, respectively. pScaf-10080.1 and the insert amplicon were both digested with KpnI and BglII for 15 min and then ligated overnight at 4°C (for detailed sequences, see Supplementary File S1). The ligation product was transformed into DH5alpha *E. coli* using heat shock. Transformed *E. coli* were grown in the presence of ampicillin on LB plates overnight. Colonies were PCR screened for the insert using the same primers mentioned earlier. After sequence confirmation, a single colony was expanded for phagemid isolation.

2.3 ssDNA scaffold production

Competent F′ *E. coli* [5-alpha F'Iq, New England Biolabs (NEB)] were transformed with the sequence validated phagemid and plated. A single isolated colony was inoculated in 20 ml of low sodium (3 g/l) LB media with 100 μg/ml ampicillin and incubated at 37°C for ~4–6 h. Then 20 μl of M13KO7 helper phage (NEB) was added to the culture and incubated at 37°C for 1 h. Kanamycin was added to the culture at a final concentration of 50 μg/ml and incubated overnight at 37°C. The culture was centrifuged at 4000 g for 20 min to pellet the bacteria cells. The cleared media was transferred to a clean 50-ml falcon tube, mixed with 0.2 volumes of 20% polyethylene glycol (PEG) with 5 mM NaCl and held at 4°C for at least 1 h and up to overnight. Samples were spun down at 7200 g for 1 h. The resulting pellet was resuspended in 800 μl Tris-buffered saline and centrifuged again to remove any remaining bacteria. This cleared media was transferred to a clean tube and again incubated with 0.2 volumes of PEG solution at room temperature for 5 min. Samples were centrifuged at 25 000 g for 25 min and the pellet was resuspended in 100 μl Tris-EDTA (pH 8) and column purified using the E.Z.N.A. M13 DNA Mini Kit (Omega Biotek). The ssDNA was run on a 1% agarose gel to confirm size and purity.

3. Results and discussion

To create a larger dNAM node, we designed a rectangular origami that allows encoding of an 8 × 10 matrix with 10 nm spacing between data strands [8]. We used the software caDNAno to design

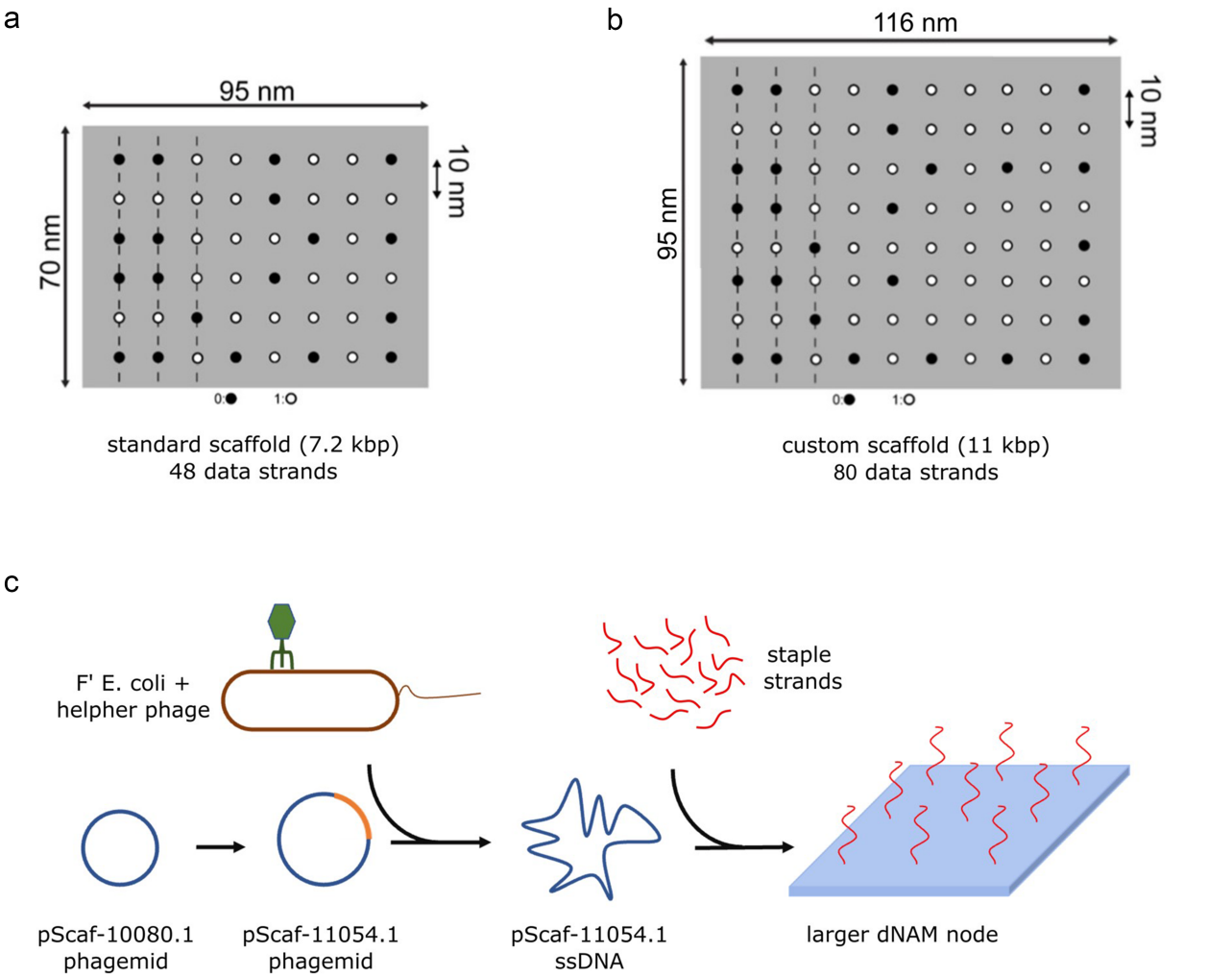


Figure 1. Larger origami design and production scheme. (a) Dimensions of the prototype 6 × 8 dNAM data matrix and (b) the new larger 8 × 10 data matrix. White circles represent sites with protruding data strands (1) and black circles represent sites without data strands (0). (c) Cloning and scaffold production plan. Restriction-based cloning adds 974 bp (bold, orange) to pScaf-10080.1 to create a dsDNA phagemid termed pScaf-11054.1. The phagemid is used to transform *E. coli*, and its co-infection with helper phage (green) results in single-stranded scaffold DNA. The custom scaffold DNA is then combined with ssDNA staple strands to synthesize a larger dNAM node that spatially orients a larger number of protruding DNA data strands (squiggly line, red).

a rectangle composed of 32 helices, each 352 bp long [14]. The resulting rectangle was ~116 nm × 95 nm (Fig. 1b) and would require a ssDNA scaffold molecule that is 11 054 nt long as well as 352 DNA oligonucleotide staple strands, each 30–32 nt in length. To position data strands on the surface of the origami, specific staple strands were extended with the sequence 5′-TTGGGAGGA-3′, which has a spacer of two thymidine nucleotides (TT) followed by a sequence that is complementary to the fluorescent imager probe for DNA-PAINT imaging.

To enable synthesis of the 11 054-nt ssDNA scaffold, we first designed and cloned a dsDNA plasmid with M13 phage origins (phagemid) that can express the single-stranded scaffold in *E. coli*. We started with a phagemid, previously reported by others, that was engineered to contain an M13 origin as well as an ssDNA terminator sequence [15]. This phagemid produces ssDNA from the sequence between the origin and terminator when transformed into *E. coli* and upon the addition of M13KO7 helper phage. The genes of the helper phage also package and export the ssDNA out from the bacteria into the media where it can be isolated using common PEG precipitation and DNA extraction protocols (see the 'Materials and methods' section). We started with a phagemid that produces a 10 080-nt scaffold (pScaf-10080.1, Addgene plasmid #111410) and added 982 bp from a second phagemid (pScaf-3024.1, Addgene plasmid #111404), with the sequence taken from between the M13 origin and terminator sequences (Supplementary Fig. S2). We PCR amplified the 982 bp portion of pScaf-3024.1 with primers that added KpnI and BglII restriction sites to the 5′ and 3′ ends, respectively. KpnI and BglII restriction sites are located 8 nt apart on the pScaf-10080.1 plasmid (Supplementary Fig. S2). Both plasmid and insert amplicon were digested with KpnI and BglII and then ligated together. DH5alpha *E. coli* were transformed with the ligation product and plated on ampicillin LB plates to select for the resistance marker in the phagemid. Six colonies were PCR screened for the 982 bp insert, and two colonies were identified as having the correct sized insert. One of these colonies with the insert was grown to produce more phagemid. We named this isolated larger phagemid pScaf-11054.1. Gel electrophoresis confirmed that the isolated phagemid was larger than the original pScaf-10080.1 phagemid (Fig. 2). Sanger sequencing confirmed 11 305 out of 13 809 bp of the phagemid, with no

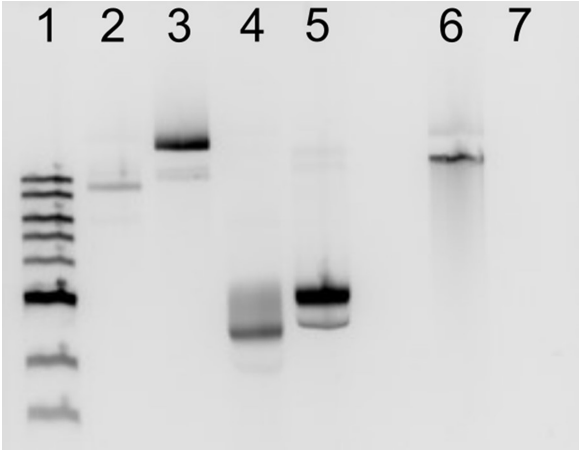


Figure 2. dsDNA phagemid and ssDNA scaffold validation by 1% agarose gel electrophoresis (1). 1 kb plus DNA ladder (NEB) (2) starting destination plasmid pScaf-10080.1 double-stranded DNA (3). Newly cloned pScaf-11054.1 double-stranded DNA (4) and single-stranded DNA scaffold produced from pScaf-10080.1 (5). Single-stranded DNA scaffold produced from pScaf-11054.1 (6). Newly cloned pScaf-11054.1 dsDNA phagemid from lane 3 treated with S1 single-stranded DNA nuclease (7). Single-stranded DNA scaffold from lane 5 treated with S1 single-stranded DNA nuclease (degraded).

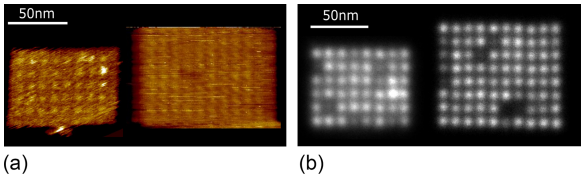


Figure 3. AFM and DNA-PAINT characterization of original and custom-produced NAM origami platform. (a) Representative AFM images of the original dNAM prototype (left) and the new larger origami using the custom pScaf-11054.1 scaffold (right). (b) Averaged DNA-PAINT representative images of the two different dNAM origami platforms. In this origami, all potential data sites have a DNA-PAINT docking site coding. Images were collected separately but are shown on the same scale.

identified mutations, including the entire 982 bp insert and both ligation junctions (Supplementary Fig. S2, Supplementary Table S1).

Next, the pScaf-11054.1 phagemid was used to produce ssDNA scaffolds. The phagemid was used to transform an F′ strain of *E. coli* which was then grown for 4–6 hours before being infected with helper phage M13KO7 followed by overnight growth to accumulate phage particles [16]. *Escherichia coli* cells were removed by centrifugation, and phage particles from the remaining bacteria-free culture were collected by PEG precipitation. The DNA was purified from proteins and PEG using silica columns. Gel electrophoresis confirmed the presence of ssDNA that was larger than the ssDNA produced by pScaf-10080.1 (Fig. 2, lanes 4 and 5). The isolated DNA was confirmed to be single stranded by S1 nuclease digestion, with the dsDNA phagemid as a negative control (Fig. 2, lanes 6 and 7). Sanger sequencing directly from the ssDNA confirmed the presence of the 982-nt insert with no mutations, ruling out contamination of the original 'parent' pScaf-10080-1 phagemid. To evaluate the yield of ssDNA from the *E. coli* cultures, we carried out three cultures in parallel and quantified an average yield of 163.1 ng/ml of starting culture. This analysis confirmed that the cloned phagemid pScaf-10080.1 produces the required ssDNA scaffold needed to synthesize the larger dNAM origami.

To functionally validate the 11 054-nt scaffold and compare it to the prototype dNAM structure, we next synthesized larger rectangular origami structures. We carried out multiple synthesis reactions, from different scaffold preparations, in order to evaluate how consistently this scaffold self-assembles into the rectangular origami. For each origami synthesis, a preparation of the ssDNA scaffold was mixed with a molar excess of 352 complementary DNA staple strands and slowly cooled from 90°C to 20°C in a thermocycler. The self-assembled origami was confirmed by agarose gel electrophoresis and then excised from the gel. The gel purified product was visualized by atomic force microscopy (AFM). We analysed five separate origami synthesis reactions, using four different preparations of the ssDNA scaffold. For each synthesis, we collected five separate AFM images with hundreds of origami structures in each. Combined, we observed 1462 individual origami, and only 66 had observable defects, indicating that 95.5% of the origami are correctly folded in the final samples (Supplementary Table S3, Supplementary Fig. S4). In addition, the origami structures had an averaged dimensions of 91 × 113 nm, which is the expected size, within the accuracy limits of AFM (Fig. 1b). To confirm the correct positioning of data strands, the full 8 × 10 matrix of data strands on the origami surface was visualized by DNA-PAINT as previously described [8]. Images were processed and averaged to produce a final image, which showed the data matrix expected from this design (Fig. 3b). We note that some data sites have lower average fluorescence intensity, which has been observed previously, and research is ongoing to investigate the cause of this lower average fluorescence at these sites. In fact, similar to other data matrix codes, such as QR codes, our dNAM data encoding scheme is designed to allow for the false negatives that can occur from lower, or missing fluorescent signals. Nevertheless, the AFM and SRM data confirm that the large ssDNA scaffold can efficiently fold into larger origami structures and that these structures effectively position protruding data strands with the nanometre precision required for dNAM data storage applications.

To demonstrate how this larger origami can be used to store data, we designed staple strands to synthesize an origami with a specific data matrix pattern. The encoding scheme used in our dNAM approach uses some of the DNA-PAINT probes for purposes other than storing data (Fig. 4a). Specifically, in this larger origami design, 3 positions are used for 'orientation' to make sure that the rectangle is decoded in the right orientation, 3 positions are used for 'indexing' which allows data spread over multiple origami to be reconstructed in the correct order, and 44 position are used for error correction ('parity' and 'checksum') which allows the correct data to be recovered even if some DNA-PAINT signals are not observed when they should be (false negatives). We designed and synthesized a dNAM origami that encoded the binary string (00100100000101010111001100001) with the appropriate orientation, indexing and error correction bits, and visualized the origami using DNA-PAINT (Fig. 4b). The resulting image demonstrates the writing (origami synthesis) and reading (DNA-PAINT) of digital data with this larger origami structure.

This new larger origami structure provides more efficient writing and reading of data in several ways. The larger origami allows more bits (80 per node) compared to the previous structure (48 bits). In addition, the number of positions dedicated to orientation and indexing do not increase for the larger origami, meaning that there is a higher ratio of 'data bits' to 'non-data bits'. This

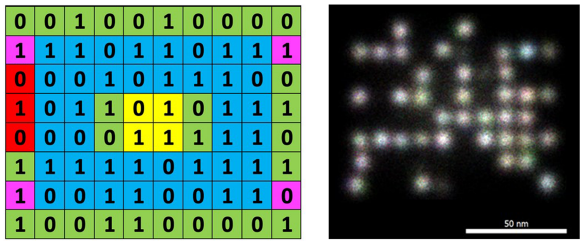


Figure 4. Demonstration of dNAM with the larger origami structure. (a) Encoding scheme for the larger origami rectangle with an 8 × 10 matrix (green = data, blue = parity, yellow = checksum, magenta = orientation, red = index). The data are read left to right, top to bottom, representing the binary string (0010010000010101011100110001). (b) An averaged DNA-PAINT image of the data node encoded as DNA origami.

efficiency means that fewer origami need to be synthesized and read by DNA-PAINT. For example, the message used in our prior proof of concept, 'Data is in our DNA!\n', required the synthesis and reading of 15 different origami structure using the smaller rectangle origami with the M13mp18 scaffold [8]. With our new larger origami structure, the same message can be encoded in only six origami. Thus, while there are only 1.67 times more DNA-PAINT sites per origami (80 sites vs. 48 sites), the larger origami is 2.5 times more efficient in data storage (6 origami vs. 15 origami). By reducing the number of origami that need to be synthesized and imaged by DNA-PAINT, the custom-sized origami demonstrated here provides a step toward more efficient dNAM data storage. As previously described, origami-based dNAM is already an improvement over existing technologies in terms of data density, stability, and energy required for production and storage. However, improvements are needed in the speed of reading and writing data, especially considering the scale of data writing and storage that is needed to meet growing demands. The results presented here were intended to advance the efficiency of data reading and writing with an eye toward scalable and efficient dNAM.

Conclusion

In conclusion, we have designed, built, and tested a custom-sized DNA scaffold for advancing the dNAM approach where data are written in DNA origami and read by DNA-PAINT. In addition, the larger scaffold presented here opens up several new possibilities for advancing dNAM [8]. For example, the demonstration that scaffolds for this large rectangular origami can be produced with standard lab approaches encourages the design of additional custom scaffolds. Future designs are expected to achieve not only a desired size, but completely custom designed sequences. The phagemid described here was produced by shuffling together DNA from two existing phagemids, but future efforts could use *de novo* sequence synthesis to produce scaffolds that use orthogonal sets of staple strands. This could allow multiple origami structures to be synthesized simultaneously in 'one pot', which would allow efficient synthesis needed for larger data sets. In addition, having multiple orthogonal scaffolds could allow specific staple strands within larger pools of DNA oligonucleotides to be selectively pulled down. This approach would allow random access of data and take advantage of recent advancements in DNA synthesis approaches. Outside of dNAM, this larger scaffold may also be useful for applications in other large-scale DNA nanotechnology designs when the spatial positioning of numerous organic or inorganic materials with nanometre precision is required.

Acknowledgments

We acknowledge AFM support from Dr Paul Davis and the Surface Science Lab at Boise State University. Undergraduate participation was supported by Dr Donna Llewellyn and the Institute for Inclusive and Transformative Scholarship at Boise State University through the Vertically Integrated Projects program.

Author contributions

S.E.K., G.D.D., W.L.H., L.P., and E.J.H. conceived and designed the project. S.E.K., M.L., A.W., B.C.B., G.M.M., and L.P. conducted experiments and analysed the data. S.E.K., L.P., and E.J.H. wrote the manuscript. All authors reviewed and revised the manuscript. T.A., W.L.H., L.P., and E.J.H. supervised the research and acquired funding.

Supplementary data

Supplementary Data is available at *SYNBIO* online.

Conflict of interest: G.M.M., G.D.D., T.A., W.L.H., L.P., and E.J.H. are listed as inventors on a pending patent US Serial No. 17/443312. This patent involves compositions and methods for encoding and retrieving data as spatial locations in nucleic acid architectures using super resolution microscopy.

Funding

This study was supported by the National Science Foundation (Award Nos 1807809 and 2227626), the Semiconductor Research Corporation, the State of Idaho through the Idaho Global Entrepreneurial Mission and Higher Education Research Council, and the Biomolecular Sciences Graduate Programs at Boise State University.

Data availability

All data generated or analysed in this paper are either included in the paper/supplementary data or can be made available upon reasonable request by the corresponding author. Plasmids are available at Addgene.

References

1. International Data Corporation. *Worldwide Global Data Sphere Forecast, 2021–2025: The World Keeps Creating More Data — Now, What Do We Do with It All?* IDC, 2021.
2. Ang J, Apalkov D, Assaderaghi F *et al. Decadal Plan for Semiconductors*. Semiconductor Research Corporation, 2021
3. Monroe J, Preston R. *Market Trends: Evolving Enterprise Data Requirements — How Much Is Not Enough?* Gartner Research, 2020
4. Fujifilm. *TCO Tool*. https://www.fujifilm.com/ca/en/business/data-storage/resources/tco-tool (13 February 2025, date last accessed).
5. Zhirnov V, Zadegan RM, Sandhu GS *et al*. Nucleic acid memory. *Nat Mater* 2016;**15**:366–70. https://doi.org/10.1038/nmat4594
6. DNA Storage Alliance. *Preserving Our Digital Legacy: An Introduction to DNA Data Storage*. 2021.
7. Doricchi A, Platnich CM, Gimpel A *et al*. Emerging approaches to DNA data storage: challenges and prospects. *ACS Nano* 2022;**16**:17552–71. https://doi.org/10.1021/acsnano.2c06748

8. Dickinson GD, Mortuza GM, Clay W *et al*. An alternative approach to nucleic acid memory. *Nat Commun* 2021;**12**:2371. https://doi.org/10.1038/s41467-021-22277-y
9. Rothemund PWK. Folding DNA to create nanoscale shapes and patterns. *Nature* 2006;**440**:297–302. https://doi.org/10.1038/nature04586
10. Schnitzbauer J, Strauss MT, Schlichthaerle T *et al*. Super-resolution microscopy with DNA-PAINT. *Nat Protoc* 2017;**12**:1198–228. https://doi.org/10.1038/nprot.2017.024
11. Kick B, Praetorius F, Dietz H *et al*. Efficient production of single-stranded phage DNA as scaffolds for DNA origami. *Nano Lett* 2015;**15**:4672–76. https://doi.org/10.1021/acs.nanolett.5b01461
12. Marchi AN, Saaem I, Vogen BN *et al*. Toward larger DNA origami. *Nano Lett* 2014;**14**:5740–47. https://doi.org/10.1021/nl502626s
13. Rafat AA, Pirzer T, Scheible MB *et al*. Surface-assisted large-scale ordering of DNA origami tiles. *Angew Chem Int Ed* 2014;**53**:7665–68. https://doi.org/10.1002/anie.201403965
14. Douglas SM, Marblestone AH, Teerapittayanon S *et al*. Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res* 2009;**37**:5001–6. https://doi.org/10.1093/nar/gkp436
15. Nafisi PM, Aksel T, Douglas SM. Construction of a novel phagemid to produce custom DNA origami scaffolds. *Synth Biol* 2018;**3**:ysy015. https://doi.org/10.1093/synbio/ysy015
16. Vieira J, Messing J. Production of single-stranded plasmid DNA. In: Wu R, Grossman L, Moldave K. *Recombinant DNA Methodology*. San Diego: Academic Press, *Selected Methods in Enzymology*, 1989, 225–233. https://doi.org/10.1016/B978-0-12-765560-4.50017-4

Synthetic Biology, 2025, 10(1), ysaf008 , DOI: https://doi.org/10.1093/synbio/ysaf008, Advance Access Publication 7 April 2025, Short Communication
Submitted: 3 April 2025; **Received (in revised form)**: 14 January 2025; **Accepted**: 6 April 2025