SCIENTIFIC REPORTS

Bioproduction of pure, kilobase-scale single-stranded DNA

Tyson R. Shepherd, Rebecca R. Du, Hellen Huang, Eike-Christian Wamhoff & Mark Bathe

Scalable production of kilobase single-stranded DNA (ssDNA) with sequence control has applications in therapeutics, gene synthesis and sequencing, scaffolded DNA origami, and archival DNA memory storage. Biological production of circular ssDNA (cssDNA) using M13 addresses these needs at low cost. However, one unmet goal is to minimize the essential protein coding regions of the exported DNA while maintaining its infectivity and production purity to produce sequences less than 3,000 nt in length, relevant to therapeutic and materials science applications. Toward this end, synthetic miniphage with inserts of custom sequence and size offers scalable, low-cost synthesis of cssDNA at milligram and higher scales. Here, we optimize growth conditions using an *E. coli* helper strain combined with a miniphage genome carrying only an f1 origin and a β-lactamase-encoding (*bla*) antibiotic resistance gene, enabling isolation of pure cssDNA with a minimum sequence genomic length of 1,676 nt, without requiring additional purification from contaminating DNA. Low-cost scalability of isogenic, custom-length cssDNA is demonstrated for a sequence of 2,520 nt using a bioreactor, purified with low endotoxin levels (<5 E.U./ml). We apply these exonuclease-resistant cssDNAs to the self-assembly of wireframe DNA origami objects and to encode digital information on the miniphage genome for biological amplification.

Kilobase-length, single-stranded DNA (ssDNA) is essential to numerous biotechnological applications including sequencing[1], cloning[2], homology directed repair templating for gene editing[3], DNA-based digital information storage[4,5], and scaffolded DNA origami[6–9]. Specifically, scaffolded DNA origami enables the fabrication of custom structured nanoscale objects with application to nanoscale lithography[10,11], light harvesting and nanoscale energy transport[12–15], metal nanoparticle casting[16], and therapeutic delivery[17,18]. In this approach, a long ssDNA scaffold is folded via self-assembly into arbitrary, user-specified shapes by slow annealing in the presence of complementary short oligonucleotide "staples". These staples are designed using Watson-Crick base-pair complementarity to the scaffold, forcing sequences that are far apart in sequence space to be close in physical space. Fully automated, top-down computational design of scaffolded DNA origami nanostructures has now been enabled by sequence design algorithms in both 2D and 3D[19–23], enabling the democratization of otherwise complex scaffolded DNA origami design that previously excluded non-experts. DNA origami generated from these algorithms with sizes from 10–50-nm typically require scaffold lengths of 1–3 kb. However, therapeutic and materials science applications that require large-scale, low-cost scaffold with custom length and sequence requirements are still hindered by limitations in production of circular, isogenic scaffolds on the 1–3 kb scale[24–28].

The most common low-cost source of native cssDNA for scaffolded DNA origami is the 7,249 base single-stranded M13mp18[29] phage genome, which has allowed production of up to 410 mg of cssDNA per liter of *E. coli* growth through fed-batch fermentation[30]. However, because therapeutic and materials science applications of DNA origami often require exact-size scaffolds less than 3,000 nt, scalable production of custom scaffold lengths are of value. To achieve custom sequence design of mini-scaffold DNA, helper plasmid systems are employed where the M13 coding sequences are sub-cloned onto a double-stranded, low-copy number vector that is co-transformed with a phagemid containing an ssDNA origin of replication (e.g., f1 origin) that allows for the synthesis and packaging of ssDNA. The most commonly used helper plasmid system is M13KO7[31], which maintains a packaging sequence, albeit with a mutated packaging signal to reduce the packaging frequency. This system has shown utility in phage display[32–34], and has also been applied to produce phagemids that encode either a 1,983-nt or 2,404-nt sequence, both containing a cssDNA f1 origin, a dsDNA pUC origin, and an ampicillin selection marker[24,28]. However, these phagemid cssDNAs were contaminated by other DNAs, both from the dsDNA phagemid and M13KO7[24]. Importantly, isolation of the target ssDNA from these DNA impurities would

Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA. Correspondence and requests for materials should be addressed to T.R.S. (email: trsheph@mit.edu) or M.B. (email: mark.bathe@mit.edu)

require subsequent purification steps for any further scale-up or would introduce background sequence contamination of the helper plasmid and dsDNA and consequently lower yields in applications to DNA origami folding.

To overcome these limitations, the ssDNA origin of replication and packaging signal was entirely removed from a helper plasmid (*E. coli* str. M13cp)[35], thereby enabling the biological production of isogenic cssDNA without DNA impurities. Leveraging this advance, one innovative approach to achieving custom bacterial scaffolds was recently implemented using a modified M13 origin of replication, but scalability was not reported and a lack of a selection marker in the produced material indicated that scalability might be challenging[25]. Another recent novel approach was the self-excision of ssDNA from phage DNA for scaffolds less than 3,200 nt[26]. However, the produced strands are linear and, without additional chemical stabilization[36], would be prone to exonuclease degradation[37]. Additionally, the excision process requires further purification steps to remove residual DNAzymes. In each case, optimization for pure cssDNA production[38] without genomic or plasmid contamination is still required[25,26]. Additionally, elimination of the antibiotic selection marker from both of these approaches[25,26] ultimately does not allow for subsequent re-infection without this selective control, limiting downstream biological applications where reinfection would be useful, such as for biological sequence amplification. Thus, there remains a critical need for scalable production of isogenic cssDNA at the 1–3 kb length scale that maintains replication capacity.

To address this need, we show that isogenic miniphage production of cssDNA is scalable by fermentation using the *E. coli* str. SS320 with the M13cp helper plasmid. Three miniphages were synthesized using both classic restriction and restriction-free (RF) cloning[39,40]. The miniphage presented here maintain the selection marker and origin of replication, which allows for the reinfection of the phage in culture while reducing the occurrence of contaminating dsDNA because they do not contain a double-strand origin of replication, similar to the natural M13 phage. Monitoring phage yields and growth rates of the bacteria, we identify an 8-hour timepoint after inoculation that yields maximal cssDNA production with no detectible DNA contamination. Scalable silica-column-based DNA extraction techniques from clarified media yielded 2 mg of pure cssDNA per liter of culture with final endotoxin levels similar to detergent based methods of endotoxin removal[41]. We demonstrate for the first time bioreactor-scalability of production of highly pure cssDNA of less than 3,000 nt in length, which is essential for the generation of circular scaffolds for wireframe DNA origami nanoparticles with partial sequence control[19,22,23], with additional applications to write-once, read-many archival DNA data storage.

Results

A variant of extension-overlap, restriction-free cloning[39,40] using long ssDNA (Fig. 1a) was applied for the *de novo* assembly of a miniphage genome containing only an f1 origin of replication and an ampicillin resistance selection marker (phPB52). Two kilobase-scale megaprimer ssDNAs were generated using asymmetric PCR (aPCR)[37] using 5′-phosphorylated primers: a top-strand megaprimer encoding the f1 origin sequence (427 nt)[42] and a bottom-strand megaprimer encoding an ampicillin resistance cistron (*bla*; 1,249 nt) (Figs 1b and S1). The kilobase primers were synthesized such that the two sequences contained a complementary sequence of 20 nt on each of the 5′ and 3′ ends (Fig. 1a). The two megaprimers were mixed at equimolar concentration and completed to dsDNA using PCR, followed by enzymatic ligation. The ligated plasmid (Fig. S2) was transformed into chemically competent *E. coli* str. M13cp[35] and dual selected on ampicillin and chloramphenicol with no detectable toxicity due to the miniphage, resulting in normal colony shape and size. Two out of eight colonies screened were found to be of the exact sequence desired. Liquid culture was inoculated and grown in a shaker flask, after which the culture was centrifuged to separate the phage-containing media from the bacterial pellet. Phage in the clarified media were visualized by TEM, showing the anticipated size and homogeneity (Figs 1c and S1). The cssDNA from this phage was isolated using silica column purification and showed 88% cssDNA purity according to agarose gel imaging (Fig. 1d), with an approximate yield of 0.5 mg per liter of bacterial growth, while the bacterial pellet showed helper plasmid, dsDNA intermediate phage DNA, and cssDNA.

Having generated a phage containing only the f1 origin and a resistance gene, demonstrated to be stably produced and exported to the media from the helper strain, we next sought to generate a second phage with a synthetic fragment of DNA that is orthogonal in sequence to bacterial and phage genomes. A fragment of length 844 nt was ligated between the f1 origin and the *bla* cistron using standard restriction cloning to generated a plasmid of size 2,520 nt (phPB84; Fig. S2). This plasmid was transformed into the helper strain and the produced phage was purified and its sequence verified by primer walking with Sanger sequencing (External Tables S1 and S2)

In order to obtain milligram-scale production of cssDNA with high genetic purity of the final material, we used a shaker flask setup (Fig. 2a) to vary the growth time, the *E. coli* strain, the growth media, and the media pH to determine optimal conditions. We found the highest and purest yield of cssDNA production occurred at the 8-hour timepoint after inoculation, near the end of log phase, with production falling off thereafter and the appearance of dsDNA contaminations in the media visualized at the 12-hour timepoint (Figs 2b and S3). Two strains were tested for production: DH5a F′Iq (Invitrogen) and the SS320 strain (Invitrogen). Both express the F pili and are commonly used for phage production, and each was transformed with the M13cp helper plasmid purified from *E. coli* str. M13cp. Strain SS320 showed approximately double the cssDNA yield (Figs 2c and S4) and was therefore chosen as the strain for further optimization of growth conditions. Terrific broth (TB) and 2 × yeast extract tryptone (2 × YT) media for bacterial growth and cssDNA production were both evaluated for phage growth while also monitoring dsDNA contamination using agarose gel analysis (Figs 2d and S4). Notably, TB had significant dsDNA contamination by the 8-hour timepoint (Fig. S4), and 2 × YT was therefore chosen as the optimal media for batch production. Next, we investigated the pH sensitivity of the production of phage material[38], which exhibited a three-fold increase in yield at pH 6.8 and 7.2 compared to pH 8 (Figs 2e and S4).

Having identified the optimal growth conditions in the shaker flask setup, we next identified conditions for scale-up in a batch fermenter process (Fig. 3a) using a Stedium Sartorius 5 L fermenter (Sartorius, Germany). Shaker flask conditions were transferred to the bioreactor setup including using 2 × YT media, while pH 7.0 was
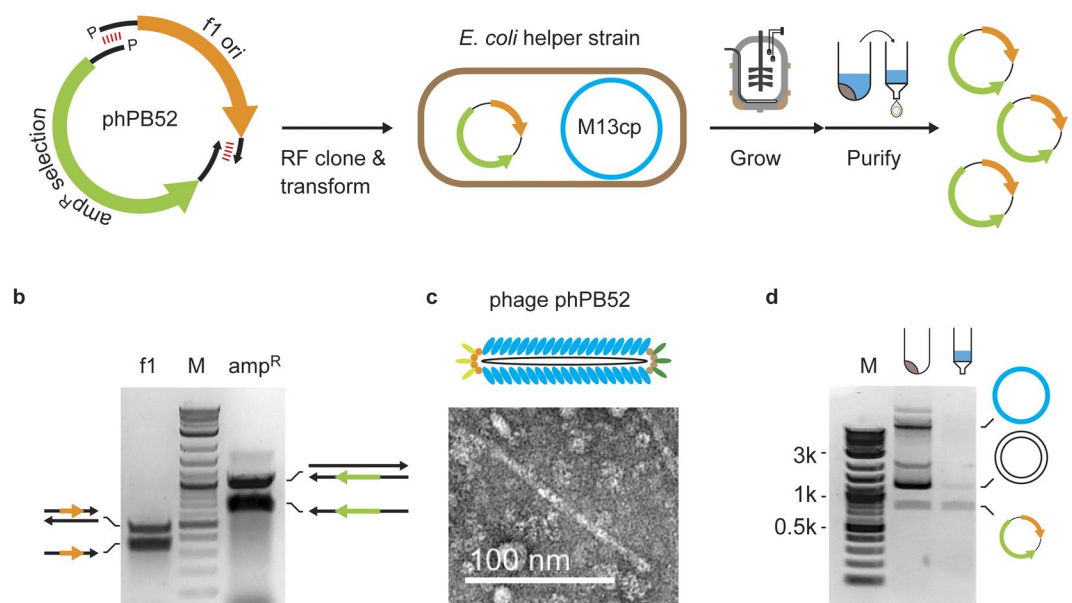
2

a Production of isogenic miniphage

Figure 1. Scalable bacteriophage production of isogenic cssDNA. (**a**) Miniphage phPB52 was assembled using restriction-free (RF) cloning was used for miniphage phPB52 assembly and transformed into *E. coli* containing the M13cp helper plasmid for production of isogenic cssDNA. (**b**) aPCR was used to generate the two ssDNA megaprimers for RF cloning encoding the f1 origin of replication (f1 ori) and the *bla* ampicillin selection marker (Selection). See Fig. S1a for uncropped color image. (**c**) Phage particles from clarified media were visualized by TEM. See Fig. S1b for uncropped image. (**d**) DNA purification from the bacterial pellet and the clarified media show mostly pure cssDNA in the media and cssDNA and dsDNA phagemid, and helper plasmid in the bacterial pellet. See Fig. S1c for uncropped color image.
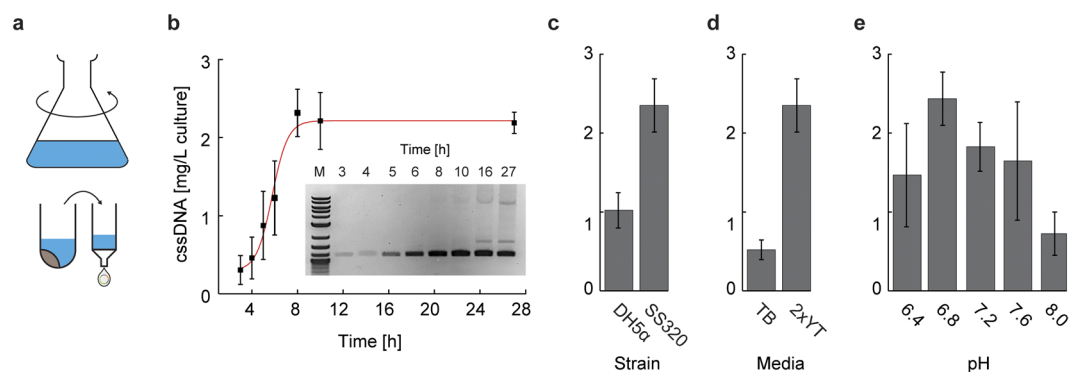

Figure 2. Shaker flask production of pure cssDNA. (**a**) Shaker flask growth of phPB84 was used to optimize conditions for phage amounts and purity. (**b**) Time-course assay of cssDNA production of phPB84, with cssDNA yield calculated by absorbance at 280 nm and purity adjusted by agarose gel band intensity, showing maximum yield and purity at the 8-hour timepoint. The 16 h time-point is from a separate culture, and therefore is not included in the plot. See Fig. S1a for uncropped color gel image. (**c**) Comparison between DH5a F′Iq and SS320 showing two-fold yield increases in the SS320 strain. (**d**) Comparison between growth media showing five-fold improved cssDNA yield in $2 \times$ YT after 8 hours of production. (**e**) Comparison of five pH values for cssDNA production, controlled by use of 100 mM HEPES-NaOH. Error bars indicate standard deviation of triplicate experiments. See Figs S3 and S3 for uncropped color triplicate measurements of all experiments.

controlled using external phosphoric acid and ammonium hydroxide. The growth curve was monitored using O.D.600 absorbance measurements and the pH and dissolved oxygen were monitored by calibrated probes. Each timepoint was additionally monitored for cssDNA and dsDNA production using agarose gel analysis (Figs 3b and S5), showing maximal cssDNA yield at the 8-hour timepoint, as with the shaker flask, with minimal contaminating dsDNA up to the 12-hour timepoint (Fig. S5). Extraction of 900 mL of media for phage purification was carried out at the 8-hour timepoint and processed using a silica-column based approach specifically designed
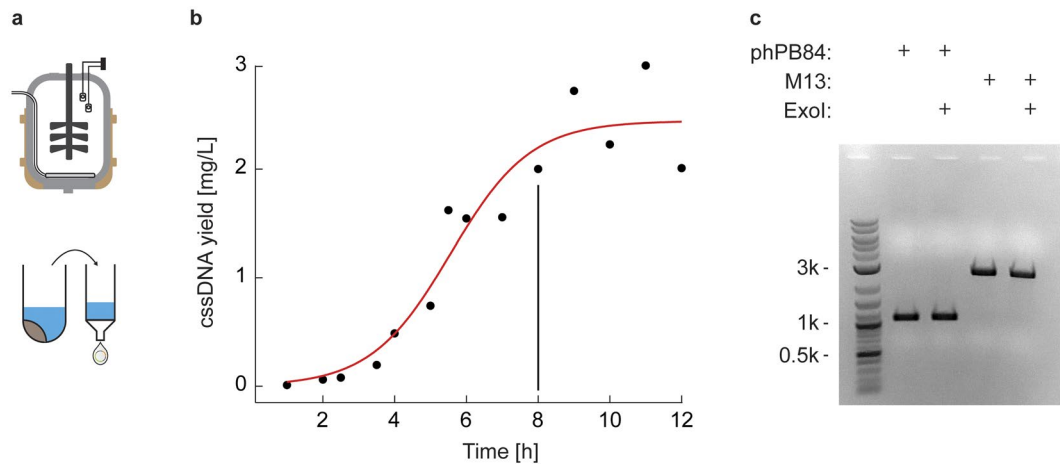
Figure 3. Batch fermenter production of pure cssDNA. (**a**) Scalable production in a stirred-tank bioreactor. (**b**) Time-course assay of cssDNA yield based on agarose gel band intensity analysis (Fig. S5), with the 8-hour timepoint used for 900 mL cssDNA purification. (**c**) Silica-column DNA purification from the PEG-precipitated phPB84 phage showed no detectible dsDNA contamination, similar in purity to commercially available M13mp18, yielding 2 mg of DNA per liter of culture at the 8-hour timepoint. Stability from exonuclease I (ExoI) degradation after 30 min coincubation indicates the ssDNA is circular.

to reduce endotoxin levels (EndoFree Megaprep Kit, Qiagen, MD). Gel band intensity analysis after kit purification showed no detectable dsDNA contamination (Fig. 3c). Sanger sequencing by primer walking verified the sequence of the phage DNA (External Tables S1 and S2). The kit-based purification yielded 2 mg of cssDNA/L of culture, matching the yield from phenol-chloroform extraction. Endotoxins were tested using a colorimetric assay (ToxinSensor Chromogenic LAL Endotoxin Assay Kit, GenScript, NJ), showing the final product yielded endotoxin levels at 1.1 ± 0.1 E.U./ml per cssDNA concentration of 10 nM, similar to endotoxin reduction by Triton-X114[41] (Fig. S6). Circularization of the produced cssDNA was verified by incubation with exonuclease I, showing no detectable degradation after 30 min (Fig. 3c).

Having implemented a method for milligram-scale production of isogenic miniphage cssDNA, we next applied the method to produce custom length single-stranded DNA scaffold with partial sequence control for application to wireframe scaffolded DNA origami (Fig. 4). We used the DAEDALUS design algorithm[19] to design a DNA-scaffolded pentagonal bipyramid with a 52-bp edge length (1,580-nt scaffold length) using the smallest phPB52 phage genome sequence (1,676 nt) and a second DNA-scaffolded pentagonal bipyramid with an 84-bp edge length (2,520-nt scaffold length) using the phPB84 phage genome sequence (2,520 nt). Notably, any DNA origami with scaffold lengths larger than 1,676 nt can have perfectly matched phage genome lengths, as exemplified in the pentagonal bipyramid with an 84-bp edge length. DNA origami object folding was characterized using agarose gel mobility shift assays and transmission electron microscopy (TEM), which confirmed monodispersed object sizes with near quantitative yield of self-assembly (Figs 4a, S7 and S8).

As an alternative application, we applied our platform to package digital information encoded in the DNA sequence for write-once, read-many archival DNA storage. Specifically, we cloned a sequence into the phPB52 variable domain that encoded a line of text ("The answer is in your memory and you need no help to give it to me. Why did you dismiss Abigail Williams?") from Act II, Scene 2 of The Crucible by Arthur Miller[43] (phCruc; Fig. 4b). While the binary representation of this encrypted text file was converted into a DNA sequence using direct nucleotide conversion (A or C representing 0 and T or G representing 1), other encryption or compression approaches could alternatively be employed. A universal forward primer, together with header information were added to the 5′ of the sequence, and an end-of-file (EOF), random slack space, and a universal reverse primer were added to the 3′ end (Fig. 4b). The sequence was optimized for single-strandedness by ensuring no regions of sequence had greater than 7 bases that were repeated or complementary to any other region of the sequence. The DNA "memory block" was cloned into the phPB52 sequence (Fig. 4b) and four-milliliter production of the phCruc phage showed 95% purity of the cssDNA as judged by agarose gel band intensities. Sanger sequencing was used to retrieve the insert sequence and decode the digital message (External Table S1).

Discussion

We applied the *E. coli* str. M13cp strain for scalable bioproduction of pure cssDNA, which has the capabilities of generating isogenic material for biotechnological applications including scaffolded DNA origami and digital information archiving and amplification, amongst other uses. The method employed here to direct purification of phage cssDNA without additional dsDNA contamination allows for new technology development in synthetic cssDNA sequence production that can be made bio-orthogonal and scalable, enabling future application to novel therapeutics and materials. Additional advances in the scaffolded DNA origami field are applicable to this strain, including the incorporation of DNAzymes[26] that would allow for greater control over the sequence and size of the produced linear ssDNA. However, in the approach used here, maintenance of the f1 origin and the selection
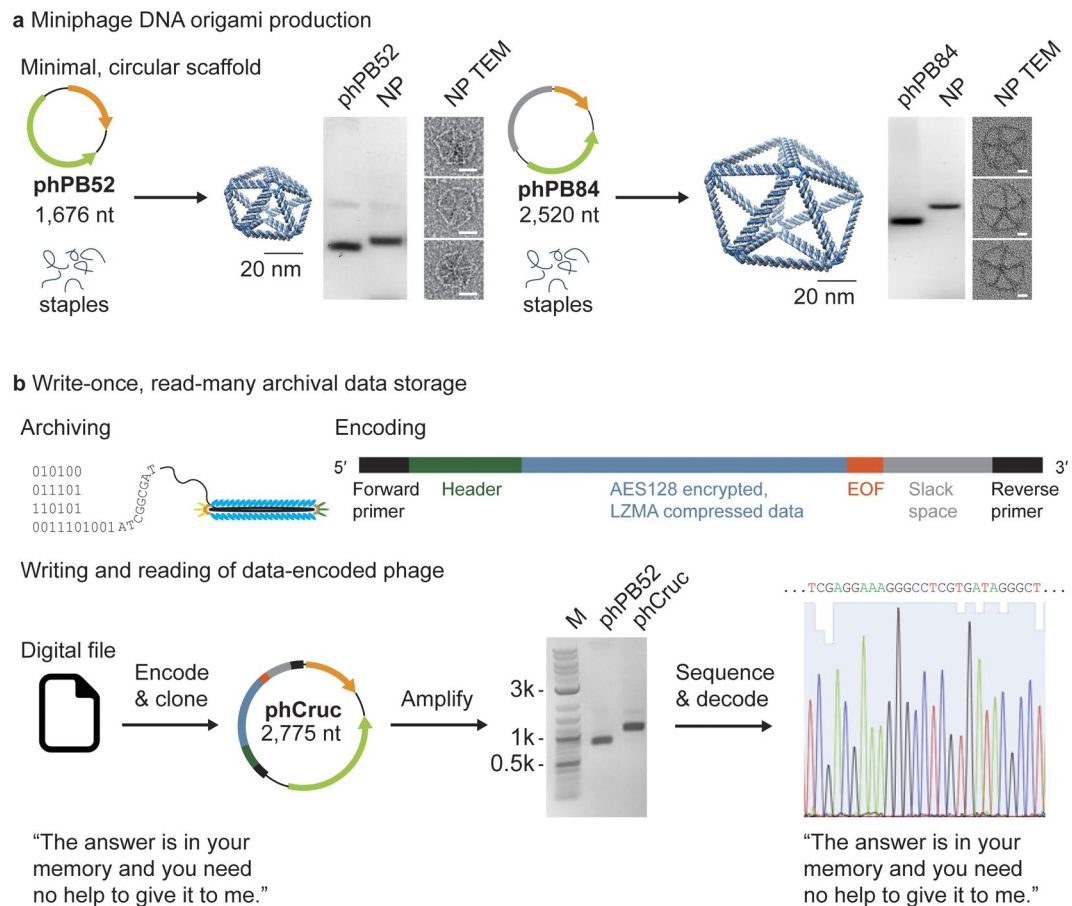
Figure 4. Applications of scalable, isogenic miniphage production. (**a**) Pentagonal bipyramids of 52-bp and 84-bp edge-lengths were folded using the phPB52 and phPB84 as scaffolds, respectively. Agarose gel shift mobility assays and TEM were used to validate the folding of the scaffold to the expected design. See Figs S7 and S8 for uncropped gel images and example full field TEM micrographs, respectively. Scale bars represent 20 nm. (**b**) Phage particles are natively protected from environmental degradation and easy to amplify by bacterial infection, and thus provide a compelling method for archival and amplification of digital information encoded in DNA. The encoding scheme shown here can generate bio-orthogonal sequences that are designed to limit secondary structure and recombination sites. DNA encoding a digital text file containing a line from The Crucible[43] (full encoded text: "The answer is in your memory and you need no help to give it to me. Why did you dismiss Abigail Williams?") was ligated to the phPB52 vector and subsequently produced and amplified in bacteria. Sanger sequencing by primer walking was used to retrieve the original digital file.

marker in the produced phage allows for reinfection across the culture, which is important for subsequent biological amplification such as needed in phage display and, here, archival information storage. Moreover, circularization blocks exonuclease activity, which may prove important for therapeutic applications[44]. In the future, improved understanding of phage biology should enable new approaches to excising specific coding sequences from M13 to generate engineered systems specifically designed for production of cssDNA.

The yields from the bioreactor approach used here were lower than wild type phage production that has been extensively optimized[26,30]. This is due in part to the loss of the native feedback control over gene expression in the phage genome[45], the use of batch fermentation as opposed to a fed-batch approach that would allow for higher cell density[30], and plasmid loss due to ampicillin selection. Interestingly, we were not able to obtain clones of kanamycin or chloramphenicol selection cistrons on the vector purely under the control of the f1 origin. This may be due to the use of their respective cognate promoters, which might be overcome by alternative single-strand-specific promoters[46]. This resistance insertion would then allow for fed-batch scale-up, leading to significantly improved yields.

Increased cssDNA production yields, together with advances in custom sequence design[25,26] and bio-orthogonality[47,48] with and without protein coding sequences, suggest that our approach is amenable to therapeutic applications in which ssDNA are used in circular[49] or linear[50] forms. In particular, scalable production of pure ssDNA at lower costs could enable yields required for therapeutic dosages of kilobase-length HDR template strands[51], a strategy that is further enabled by applying a DNAzyme approach for linearization[26]. Scalable biological production of scaffolded DNA origami now matches production amounts from solid-state DNA synthesis commonly used for staple production, so that scaffolded DNA origami nanoparticles may now be produced at

reasonable cost for mouse and higher animal therapeutic studies. Staple sequences synthesized with modifications to improve staple stability may further enhance nanoparticle lifetimes[36].

The alternative application of custom length and sequence scaffolds to encode digital information offers a write-once, read-many approach to low-cost massive archival data storage. Phage packaging is known to improve DNA stability against nuclease and chemical degradation[52], and ease of amplification in bacterial cultures makes this a intriguing method for native archival storage and biological-based information amplification, compatible with all sequencing strategies developed for M13 shotgun sequencing. Further, knowledge of phage biology and phage display offers an interesting set of possibilities for conditional amplification of phage sequences that encode specific digital information, a possible alternative or complement to the current PCR-based solutions that are being developed[53].

Materials and Methods

Plasmid assembly by single-stranded DNA.   All sequences of phage genomes (External Table S1) and primers (External Table S2) are contained in the External Supplementary Tables Excel file. The sequence of the f1 origin of replication was ordered from Integrated DNA Technologies (IDT, Inc., Coralville, IA) as a gBlock™ with 20 nt primers flanking the 5′ and 3′ sides designed to have a calculated melting temperature of 57 °C[54]. Double stranded DNA was generated by amplification of the synthetic gBlock f1 sequence with Phusion™ polymerase (New England Biolabs, Inc., Ipswitch, MA). The beta-lactamase (*bla*) ampicillin resistance gene with its promoter and terminator sequences were amplified from pUC19 using Phusion™ polymerase and 5′ and 3′ primers extended on their 5′ by the complementary pair of the f1 gBlock fragment. In each case, the PCR-amplified material was purified by ZymoClean agarose gel purification (Zymo Research, Inc., Irvine, CA) and column cleanup (Qiagen miniprep spin purification kit, Qiagen, Inc., Germany). Single-stranded DNA was generated using asymmetric production with 200 ng of purified dsDNA and 1 μM 5′-phosphorylated primer and Accustart HiFi polymerase (QuantaBio, Inc., Beverly, MA) in 1× Accustart HiFi buffer with 2 mM MgCl$_2$, and cycled 25 times, as previously described[37]. The ssDNA was gel- and column-purified. The two ssDNA products were then mixed in a 1:1 molar ratio and the ssDNA was converted to dsDNA using Phusion polymerase, column purified, and ligated using T4 DNA ligase (NEB) in 1× T4 DNA ligation buffer with 30 ng of amplified DNA incubated at room temperature overnight.

E. coli strains M13cp[35], DH5α F′Iq (Thermo Fisher, Inc., Waltham, MA), and SS320 (Lucigen, UK) were each transformed with the M13cp helper plasmid (a generous gift of Dr. Andrew Bradbury, Los Alamos National Lab) and made competent by washing log-phase grown cells in ice cold 100 mM CaCl$_2$. 20 μL of competent cells were transformed with 2 μL of phagemid DNA ligation mix. Cells were incubated on ice for 30 minutes, heat shocked at 42 °C for 45 seconds, and then put on ice. Pre-warmed SOB media was added and the cell culture was shaken at 37 °C for 1 hour. 100 μL of cells were plated evenly across a Luria-Agar (LA) media plate made with 100 μg/mL ampicillin and 15 μg/mL chloramphenicol.

Individual colonies were selected and grown in 5 mL of Terrific Broth (TB) supplemented with 1% glycerol for overnight at 37 °C. Bacteria was removed by centrifuging at 4,000 rpm for 10 minutes. Supernatant was removed and placed in a new 1.5 mL spin column and spun at 4,000 rpm for an additional 10 minutes. 1 μL of the clarified supernatant was added to 20 μL of nuclease-free water and heated to 95 °C for 5 minutes, after which 1 μL of the heated solution was added to a Phusion PCR mix containing enzyme, buffer, nucleotides, and forward and reverse primers used to generate the plasmid. Positive colonies were determined by the presence of the PCR amplicon as visualized by agarose gel, and the purified phage were sent for Sanger sequencing. Of the eight colonies chosen, two were shown to have the correct sequence. The bacterial pellet was processed to purify all containing DNA by alkaline lysis and column purification (Qiagen miniprep spin kit, Qiagen, Inc., Germany).

Purified dsDNA was PCR amplified from phPB52 to have an EcoRI and PstI nuclease sites between the *bla* resistance cistron and the f1 ori. The PCR product was purified and digested alkaline phosphatase treated and gel purified. Synthetic DNA insert encoding digital information was generated using a computational algorithm that optimizes single-stranded compatibility. The synthetic sequences were amplified from a pUC19 vector containing the sequences to have flanking EcoRI and PstI sites, and digested with EcoRI and PstI nucleases. The product was gel purified. The inserts were ligated to the phPB52 digested vector in 1×T4 DNA ligation buffer (NEB) with 30 ng of vector DNA with three molar excess of synthetic inserts, incubated at room temperature overnight, and transformed into competent helper strain *E. coli*. These generated phPB84 and phCruc phage genomes.

Synthetic phage production.   Phage producing colonies, as judged by positive PCR, gel visualization, and sequencing results, were grown overnight in 4 mL 2 × YT supplemented with 100 μg/mL ampicillin, 15 μg/mL of chloramphenicol and 5 μg/mL of tetracycline (Sigma-Aldrich, Inc.) in 15 mL culture tubes shaken at 200 RPMs at 37 °C. The following day, the cultures were diluted to an O.D.600 of 0.05 in 2 × YT supplemented with 100 μg/mL ampicillin, 15 μg/mL of chloramphenicol and 5 μg/mL of tetracycline and grown between 3 h to 27 h for time course experiments, and 8 h for media, pH, and strain optimization experiments. For pH optimization, the pH was controlled by addition of 100 mM HEPES-NaOH to the 2 × YT media. Strain-specific antibiotics were used as recommended by the manufacture. After the chosen time point, the cultures were spun down at 4,000 RPMs for 15 minutes, after which the supernatant was removed to a fresh tube and spun at 4,000 RPMs for an additional 15 minutes, and filtered using a 0.45 μm cellulose acetate filter. For gel and nanodrop quantification, 400 μL of the clarified media was lysed by addition of Qiagen Buffer P1 supplemented with Proteinase K (20 μg/mL final; Sigma) and RNase A/T1 and incubated at 37 °C for 1 h, followed by addition of Qiagen Buffer P2 and heating to 70 °C for 15 minutes, and letting return to room temperature. Qiagen Buffer N3 was then added and precipitant was centrifuged. One volume of 100% ethanol was added to the supernatant and applied to a Qiagen spin column, and purified. The purified eluate DNA concentration was determined by A280 absorbance from a NanoDrop 2000 (Thermo Fisher) for each time point and condition tested, and ran on a 1% agarose gel in 1 × Tris-Acetate-EDTA

(TAE) stained with SybrSafe (Thermo Fisher) for visualization of the product. The ssDNA purity was judged by ImageJ[55] intensity analysis and the amount of ssDNA from the time point or condition was adjusted by this purity multiplied by the total amount of DNA found from A280 absorbance.

For milligram-scale production of synthetic miniphage, a Stedium Sartorius fermenter was used for growing 5 L of culture. An overnight culture was grown in $2 \times$ YT supplemented with 100 μg/mL Ampicillin and 15 μg/mL of chloramphenicol and 5 μg/mL of tetracycline and diluted to O.D. 600 of 0.05 for inoculating 5 L of media. The growth media for the batch fermentation was also $2 \times$ YT supplemented with 100 μg/mL Ampicillin and 15 μg/mL of chloramphenicol and 5 μg/mL of tetracycline. Oxygen and pH were monitored throughout the growth, and the pH was maintained at 7.0 with phosphoric acid and ammonium hydroxide, with a constant agitation of 400 RPM. Time points were taken approximately every hour and samples were processed as above for the shaker flask. At 8 h, 900 mL of liquid culture was removed for processing. For milligram-scale purification of ssDNA, 900 mL of liquid culture bacteria was pelleted by centrifuging twice at $4,000 \times$ g for 20 min, followed by 0.45 μm cellulose acetate filtration. Phage from clarified media were precipitated by adding 6% w/v of polyethylene glycol-8000 (PEG-8000) and 3% w/v of NaCl and stirring continuously at 4 °C for 1 h. Precipitated phage were collected by centrifuging at $12,000 \times$ g for 1 h, and the PEG-8000 supernatant was removed completely, and pellet was resuspended in 30 mL of 10 mM Tris-HCl pH 8.0, 1 mM Ethylenediaminetetraacetic acid (EDTA) buffer (TE buffer). The phage was then processed using an EndoFree Maxiprep (Qiagen, Germany) column-based purification, following the manufacturer's protocol with two adjustments. First, proteinase K (20 μg/mL final) was added to EndoFree Buffer P1 and incubated at 37 °C for 1 h before addition of EndoFree Buffer P2 and incubation at 70 °C for 10 min. The lysed phage was returned to room temperature before proceeding. Second, after removal of endotoxins, 0.2 v/v of 100% ethanol was added to the clarified sample, before applying to the EndoFree Maxiprep column to increase ssDNA binding. All other steps remained the same, and the cssDNA was eluted in 1 mL of endotoxin-free TE buffer. The amount of collected DNA was judged by absorbance at A280, and the purity was judged by running on a 1% agarose gel in $1 \times$ TAE stained with ethidium bromide.

Endotoxin amounts were tested using the ToxinSensor chromogenic LAL endotoxin assay kit (GenScript, Piscataway, NJ) following the manufacturer's protocol. The cssDNA phPB84 was diluted to 10 nM in endotoxin-free water, with absorbance read for each measurement on a standard curve and the 10 nM sample on an Evolution 220 UV/Vis spectrophotometer (Thermo Fisher). Stability from exonuclease I degradation was tested by incubating cssDNA phPB84 with exonuclease I in $1 \times$ exonuclease buffer (NEB) at 37 °C for 30 min. The reaction was quenched by incubating the reaction at 80 °C for 15 min, and was subsequently ran on a 1% agarose gel in $1 \times$ TAE stained with ethidium bromide.

Agarose gels were either cast as 1% gels using SeaKem agarose stained with SybrSafe or were purchased as precast Reliant 1% Gold agarose stained with ethidium bromide (SeaKem). Gels were imaged under a blue light and the images were then made to black-and-white, and color inverted to show a black band using Adobe Photoshop CC2018. Original gel images can be found in the associated Supplementary Information.

DNA origami assembly. DNA purified from phages phPB52 and phPB84 were used to fold a pentagonal bipyramid with edge length 52 base pairs and 84 base pairs, respectively. Staples for each object were generated from the automated scaffold routing and staple design software DAEDALUS[19]. Staples were synthesized by IDT, and listed in External Tables S3 and S4. To fold the nanoparticles, 20 nM of bacterially-produced and purified scaffold was incubated with 20-molar excess of staples in $1 \times$ TAE buffer with 12 mM MgCl$_2$. The objects were annealed over 13 hours from 95 °C to 24 °C as previously described[19], and the folded particles were run on 1% agarose gel in $1 \times$ TAE buffer with 12 mM MgCl$_2$ with the respective cssDNA scaffolds for reference. The folded nanoparticles were purified using a 100 kDA MWCO spin concentrator (Amicon) for a total of five 5-fold buffer exchanges as purification for TEM.

DNA data encoding scheme. Kilobase length-scale synthesis of DNA sequences are optimized for single-strandedness by ensuring limited self-complementarity and repeat sequences, as well as reducing mononucleotide repeats[37]. Digital information can be encoded to satisfy these requirements by pseudo-randomizing the bitstream through encryption and having bit-to-base be a one-to-two encoding scheme. This was implemented using a python script that converts a digital file to a DNA sequence that satisfies constraints for single-strandedness by choosing the sequence based on the sequence having no problematic sequences of wordsize greater than 7 returned by BLASTn. The digital file bitstream is encrypted using AES with block cipher mode with a randomly generated password ("ry%Tr*>2Y><NFv5aqAEhU@Q046Cy$n92") and a randomly generated 16 nt DNA sequence initialization vector ("TAATTTACTTATTCTC"). The encoded bitstream is then translated to a DNA sequences with 0 represented randomly by an A or C and 1 represented randomly by a G or T. No homopolymers greater than 5 nucleotides are allowed and are swapped to the other choice base as required. The converted sequence is then a concatenate of a master forward primer sequence (CTTGGGTGGAGAGGCTATTC), a file type identifier (GTTTAAGGTCACATCGCATG), the initialization vector, a four-nt base-4 (T: 0, G: 1, A: 2, C: 3) memory page (TGTT), an eight-nt base-4 digital file size (GTTGAGCC), the bitstream data, an end-of-file sequence (GTACTAGTCGACGCGTGGCC) and randomly generated slack space to a user specified DNA block length, and a master reverse primer sequence (GATCTCCTGTCATCTCACCT). The slack space is randomly generated, as are the individual bit-to-base choices, and thus the bases are iteratively swapped as needed to satisfy the BLASTn wordsize constraint.

To reconstruct the file, the extraneous header and footer data is stripped from the digital encoded sequence and the sequence is converted back to bitstream data by direct conversion of A or C to 0 and G or T to 1. Applying the Python 3.4 PyCrypto module's AES function in block cipher mode with the 16-nt sequence initialization vector and password given above to the bitstream data allows for the retrieval of the original text file containing the line from The Crucible ("The answer is in your memory and you need no help to give it to me. Why did you dismiss Abigail Williams?"). We have made the source of this Python decoding algorithm freely available on GitHub at https://github.com/lcbb/ssDNA-memory.

Transmission electron microscopy. The structured DNA pentagonal bipyramid with 52-base-pair and 84-base-pair edge length assembled using the phage-produced scaffold was visualized by transmission electron microscopy (TEM). A volume of 200 μL of folded reaction was purified from excess staples and buffer exchanged into 20 mM Tris-HCl pH 8.0 and 8 mM MgCl$_2$ using a 100 kDa MWCO spin concentrator (Amicon, Merck Millipore, Billerica, MA). The concentration was subsequently adjusted to 5 nM. Carbon film with copper grids (CF200H-CU; Electron Microscopy Sciences Inc., Hatfield, PA) were glow discharged and the sample was applied for 60 seconds. The sample was then blotted from the grid using Whatman 42 ashless paper, and the grid was placed on drop of freshly prepared 1% uranyl-formate with 5 mM NaOH for 10 s[56]. Remaining stain was wicked away using Whatman 42 paper and dried before imaging. The grid was imaged on a Technai FEI with a Gatan camera.

Data Availability

Each miniphage described within is available by request to the authors. A Python script to encode and decode digital data is available as open source under GPLv3 license on GitHub at https://github.com/lcbb/ssDNA-memory.

References

1. Messing, J., Crea, R. & Seeburg, P. H. A system for shotgun DNA sequencing. *Nucleic Acids Res* **9**, 309–321 (1981).
2. Zoller, M. J. & Smith, M. Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment of DNA. *Nucleic Acids Res* **10**, 6487–6500 (1982).
3. Chen, F. *et al*. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods* **8**, 753–755, https://doi.org/10.1038/nmeth.1653 (2011).
4. Church, G. M., Gao, Y. & Kosuri, S. Next-generation digital information storage in DNA. *Science* **337**, 1628, https://doi.org/10.1126/science.1226355 (2012).
5. Goldman, N. *et al*. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **494**, 77–80, https://doi.org/10.1038/nature11875 (2013).
6. Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325**, 725–730, https://doi.org/10.1126/science.1174251 (2009).
7. Douglas, S. M. *et al*. Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418, https://doi.org/10.1038/nature08016 (2009).
8. Rothemund, P. W. & Folding, D. N. A. to create nanoscale shapes and patterns. *Nature* **440**, 297–302, https://doi.org/10.1038/nature04586 (2006).
9. Sharma, J. *et al*. Control of self-assembly of DNA tubules through integration of gold nanoparticles. *Science* **323**, 112–116, https://doi.org/10.1126/science.1165831 (2009).
10. Diagne, C. T., Brun, C., Gasparutto, D., Baillin, X. & Tiron, R. DNA Origami Mask for Sub-Ten-Nanometer Lithography. *ACS Nano* **10**, 6458–6463, https://doi.org/10.1021/acsnano.6b00413 (2016).
11. Surwade, S. P., Zhao, S. & Liu, H. Molecular lithography through DNA-mediated etching and masking of SiO2. *J Am Chem Soc* **133**, 11868–11871, https://doi.org/10.1021/ja2038886 (2011).
12. Dutta, P. K. *et al*. DNA-Directed Artificial Light-Harvesting Antenna. *Journal of the American Chemical Society* **133**, 11985–11993, https://doi.org/10.1021/ja1115138 (2011).
13. Hemmig, E. A. *et al*. Programming Light-Harvesting Efficiency Using DNA Origami. *Nano Letters* **16**, 2369–2374, https://doi.org/10.1021/acs.nanolett.5b05139 (2016).
14. Banal, J. L., Kondo, T., Veneziano, R., Bathe, M. & Schlau-Cohen, G. S. Photophysics of J-Aggregate-Mediated Energy Transfer on DNA. *J Phys Chem Lett* **8**, 5827–5833, https://doi.org/10.1021/acs.jpclett.7b01898 (2017).
15. Boulais, E. *et al*. Programmed coherent coupling in a synthetic DNA-based excitonic circuit. *Nat Mater* **17**, 159–166, https://doi.org/10.1038/nmat5033 (2018).
16. Sun, W. *et al*. Casting inorganic structures with DNA molds. *Science* **346**, 1258361, https://doi.org/10.1126/science.1258361 (2014).
17. Douglas, S. M., Bachelet, I. & Church, G. M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831–834, https://doi.org/10.1126/science.1214081 (2012).
18. Zhao, Y. X. *et al*. DNA origami delivery system for cancer therapy with tunable release properties. *ACS Nano* **6**, 8684–8691, https://doi.org/10.1021/nn3022662 (2012).
19. Veneziano, R. *et al*. Designer nanoscale DNA assemblies programmed from the top down. *Science* **352**, 1534, https://doi.org/10.1126/science.aaf4388 (2016).
20. Benson, E. *et al*. DNA rendering of polyhedral meshes at the nanoscale. *Nature* **523**, 441–444, https://doi.org/10.1038/nature14586 (2015).
21. Douglas, S. M. *et al*. Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res* **37**, 5001–5006, https://doi.org/10.1093/nar/gkp436 (2009).
22. Jun, H. *et al*. Autonomously designed free-form 2D DNA origami. *Sci Adv* **5**, eaav0655, https://doi.org/10.1126/sciadv.aav0655 (2019).
23. Jun, H. *et al*. Automated Sequence Design of 3D Polyhedral Wireframe DNA Origami with Honeycomb Edges. *ACS Nano*, https://doi.org/10.1021/acsnano.1028b08671 (2019).
24. Brown, S. *et al*. An easy-to-prepare mini-scaffold for DNA origami. *Nanoscale* **7**, 16621–16624, https://doi.org/10.1039/c5nr04921k (2015).
25. Nafisi, P. M., Aksel, T. & Douglas, S. M. Construction of a novel phagemid to produce custom DNA origami scaffolds. *Synthetic Biology* **3**, ysy015 (2018).
26. Praetorius, F. *et al*. Biotechnological mass production of DNA origami. *Nature* **552**, 84–87, https://doi.org/10.1038/nature24650 (2017).
27. Qi, X. *et al*. Programming molecular topologies from single-stranded nucleic acids. *Nat Commun* **9**, 4579, https://doi.org/10.1038/s41467-018-07039-7 (2018).
28. Zadegan, R. M. *et al*. Construction of a 4 zeptoliters switchable 3D DNA box origami. *ACS Nano* **6**, 10050–10053, https://doi.org/10.1021/nn303767b (2012).
29. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119 (1985).
30. Kick, B., Praetorius, F., Dietz, H. & Weuster-Botz, D. Efficient Production of Single-Stranded Phage DNA as Scaffolds for DNA Origami. *Nano Lett* **15**, 4672–4676, https://doi.org/10.1021/acs.nanolett.5b01461 (2015).
31. Vieira, J. & Messing, J. Production of single-stranded plasmid DNA. *Methods Enzymol* **153**, 3–11 (1987).
32. Pasqualini, R. & Ruoslahti, E. Organ targeting *in vivo* using phage display peptide libraries. *Nature* **380**, 364–366, https://doi.org/10.1038/380364a0 (1996).
33. Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. Making antibodies by phage display technology. *Annu Rev Immunol* **12**, 433–455, https://doi.org/10.1146/annurev.iy.12.040194.002245 (1994).

34. Ferrara, F., Kim, C. Y., Naranjo, L. A. & Bradbury, A. R. Large scale production of phage antibody libraries using a bioreactor. *MAbs* **7**, 26–31, https://doi.org/10.4161/19420862.2015.989034 (2015).
35. Chasteen, L., Ayriss, J., Pavlik, P. & Bradbury, A. R. Eliminating helper phage from phage display. *Nucleic Acids Res* **34**, e145, https://doi.org/10.1093/nar/gkl772 (2006).
36. Conway, J. W., McLaughlin, C. K., Castor, K. J. & Sleiman, H. DNA nanostructure serum stability: greater than the sum of its parts. *Chem Commun (Camb)* **49**, 1172–1174, https://doi.org/10.1039/c2cc37556g (2013).
37. Veneziano, R. *et al. In vitro* synthesis of gene-length single-stranded DNA. *Sci Rep* **8**, 6548, https://doi.org/10.1038/s41598-018-24677-5 (2018).
38. Reddy, P. & McKenney, K. Improved method for the production of M13 phage and single-stranded DNA for DNA sequencing. *Biotechniques* **20**(854-856), 858–860, https://doi.org/10.2144/96205st05 (1996).
39. Bryksin, A. V. & Matsumura, I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *Biotechniques* **48**, 463–465, https://doi.org/10.2144/000113418 (2010).
40. van den Ent, F. & Lowe, J. RF cloning: a restriction-free method for inserting target genes into plasmids. *J Biochem Biophys Methods* **67**, 67–74, https://doi.org/10.1016/j.jbbm.2005.12.008 (2006).
41. Hahn, J., Wickham, S. F., Shih, W. M. & Perrault, S. D. Addressing the instability of DNA nanostructures in tissue culture. *ACS Nano* **8**, 8765–8775, https://doi.org/10.1021/nn503513p (2014).
42. Dotto, G. P., Horiuchi, K. & Zinder, N. D. The functional origin of bacteriophage f1 DNA replication. Its signals and domains. *J Mol Biol* **172**, 507–521 (1984).
43. Miller, A. The Crucible. Act II Scene **2** (1953).
44. Lovett, S. T. The DNA Exonucleases of Escherichia coli. *EcoSal Plus* 4, https://doi.org/10.1128/ecosalplus.4.4.7 (2011).
45. Smeal, S. W., Schmitt, M. A., Pereira, R. R., Prasad, A. & Fisk, J. D. Simulation of the M13 life cycle II: Investigation of the control mechanisms of M13 infection and establishment of the carrier state. *Virology* **500**, 275–284, https://doi.org/10.1016/j.virol.2016.08.015 (2017).
46. Masai, H. & Arai, K. Frpo: a novel single-stranded DNA promoter for transcription and for primer RNA synthesis of DNA replication. *Cell* **89**, 897–907 (1997).
47. Kozyra, J. *et al.* Designing Uniquely Addressable Bio-orthogonal Synthetic Scaffolds for DNA and RNA Origami. *ACS Synth Biol* **6**, 1140–1149, https://doi.org/10.1021/acssynbio.6b00271 (2017).
48. Rovner, A. J. *et al.* Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518**, 89–93, https://doi.org/10.1038/nature14095 (2015).
49. Seidl, C. I. & Ryan, K. Circular single-stranded synthetic DNA delivery vectors for microRNA. *PLoS One* **6**, e16925, https://doi.org/10.1371/journal.pone.0016925 (2011).
50. Davis, L. & Maizels, N. Two Distinct Pathways Support Gene Correction by Single-Stranded Donors at DNA Nicks. *Cell Rep* **17**, 1872–1881, https://doi.org/10.1016/j.celrep.2016.10.049 (2016).
51. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* **34**, 339–344, https://doi.org/10.1038/nbt.3481 (2016).
52. Clark, J. R. & March, J. B. Bacteriophages and biotechnology: vaccines, gene therapy and antibacterials. *Trends Biotechnol* **24**, 212–218, https://doi.org/10.1016/j.tibtech.2006.03.003 (2006).
53. Yazdi, S., Yuan, Y., Ma, J., Zhao, H. & Milenkovic, O. A rewritable, random-access DNA-based storage system. *Sci. Rep*. **5**, 14138, https://doi.org/10.1038/srep14138 (2015).
54. SantaLucia, J. Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci USA* **95**, 1460–1465 (1998).
55. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675 (2012).
56. Castro, C. E. *et al.* A primer to scaffolded DNA origami. *Nature Methods* **8**, 221–229, https://doi.org/10.1038/nmeth.1570 (2011).

Acknowledgements

Author Contributions

T.R.S. and M.B. initiated the study; T.R.S., H.H., R.R.D., and M.B. designed and analyzed growth experiments, T.R.S. and E.W. collected structural characterization data, T.R.S., R.R.D. and M.B. wrote the manuscript; All authors read and edited the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-42665-1.

Competing Interests: T.R.S., R.R.D. and M.B. are coinventors on a patent pending (62/584,664) for some of the methods disclosed herein.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.