Enzymatic production of 'monoclonal stoichiometric' single-stranded DNA oligonucleotides

Cosimo Ducani[1], Corinna Kaul[1], Martin Moche[2], William M Shih[3–5] & Björn Högberg[1]

Single-stranded oligonucleotides are important as research tools, as diagnostic probes, in gene therapy and in DNA nanotechnology. Oligonucleotides are typically produced via solid-phase synthesis, using polymer chemistries that are limited relative to what biological systems produce. The number of errors in synthetic DNA increases with oligonucleotide length, and the resulting diversity of sequences can be a problem. Here we present the 'monoclonal stoichiometric' (MOSIC) method for enzyme-mediated production of DNA oligonucleotides. We amplified oligonucleotides from clonal templates derived from single bacterial colonies and then digested cutter hairpins in the products, which released pools of oligonucleotides with precisely controlled relative stoichiometric ratios. We prepared 14–378-nucleotide MOSIC oligonucleotides either by *in vitro* rolling-circle amplification or by amplification of phagemid DNA in *Escherichia coli*. Analyses of the formation of a DNA crystal and folding of DNA nanostructures confirmed the scalability, purity and stoichiometry of the produced oligonucleotides.

Since the first description of the solid-phase synthesis technique[1], it has been the method of choice for the rapid synthesis of DNA oligomers. Many recent advances in genetics and molecular biology, including, for example, the recent massive sequencing efforts[2], would not have been possible without DNA oligonucleotides, or oligodeoxyribonucleotides (ODNs), provided by the solid-phase synthesis method.

The length and accuracy of polymers of a precise sequence that can be produced by current polymer chemistry pales in comparison to what is produced in nature[3]. The number of errors in synthetic DNA increases with the number of nucleotides, and purities of only 70% are common in commercial ODNs comprising 51 bases[4]. Purification after synthesis should in principle help lower the error rate, but low purities are reported even for ODNs purified by high-performance liquid chromatography (HPLC)[5]. When very monodisperse ODNs are needed for experiments, the standard purification method is usually PAGE, which is labor-intensive, leading to high cost for such ODNs. Moreover, when PAGE-purified ODNs are used in experiments, negative effects attributable to impurities in ODNs are often observed[6–8], indicating that PAGE purification, at best, provides only an enrichment of the desired product.

Given the problems with the quality of synthetic ODNs, scientists have tried alternate ways of producing them. The use of enzymes to produce ODNs is an attractive alternative because enzymes have low error rates and can in general copy any given DNA sequence with high reliability. For example, PCR has been used to produce DNA oligomers[9–12]. One problem, however, is that PCR relies on synthetic ODNs for the primer sequence, and every PCR product includes a synthetic ODN at its 5′ end. Also, the primer needs to be of a certain length, usually >20 nucleotides (nt), and that extra sequence can create problems with purification, especially when the length of produced ODNs is similar to that of the primers. In addition, the PCR end-product is double-stranded, so extra steps are needed to create single-stranded ODNs required for most applications: primers must be chemically labeled for extraction and for elimination of unwanted strands. The combination of removing primers and complementary strands adds costs and complexity, and can increase the possibility of introducing errors.

To overcome some of these difficulties, methods using rolling-circle amplification (RCA) strategies have been tested[10,13]. These methods rely on the ligation of small ODN circles that subsequently act as templates for amplification by a strand-displacing phi29 polymerase. The resulting long single strand contains the complement to the circle sequence repeated many times, a so-called tandem repeat. This tandem repeat can be cut into monomers by restriction enzymes to release the desired ODN. Both PCR-based and RCA-based methods outlined above present ways to replicate synthetic ODNs. Errors in the original pool of ODNs will, however, be propagated to the final product and no substantial enrichment of sequence purity is achieved.

Here we present our MOSIC method for enzyme-mediated production of oligonucleotides that can be used to produce single-stranded ODNs directly from sequence-verified templates derived from single colonies of bacteria.

[1]Swedish Medical Nanoscience Center, Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden. [2]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [3]Wyss Institute for Biologically Inspired Engineering at Harvard, Boston, Massachusetts, USA. [4]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [5]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA. Correspondence should be addressed to B.H. (bjorn.hogberg@ki.se).

RECEIVED 18 FEBRUARY; ACCEPTED 26 APRIL; PUBLISHED ONLINE 2 JUNE 2013; DOI:10.1038/NMETH.2503

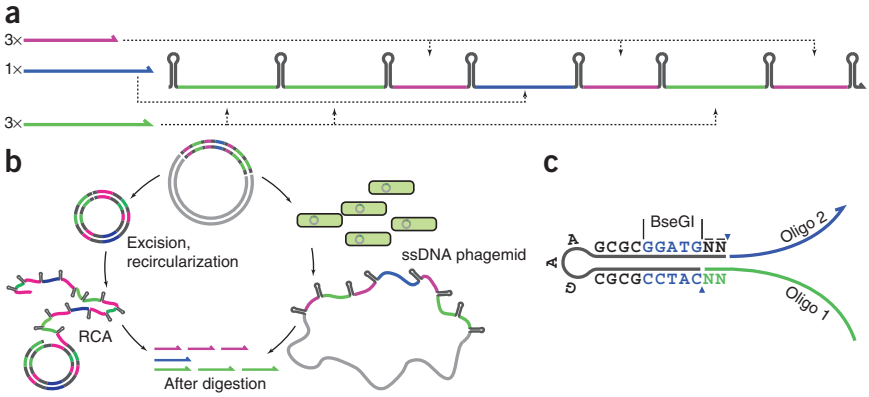


Figure 1 | Outline of the MOSIC ODN production method. (**a**) Schematic of oligonucleotides (left) with their relative stoichiometry (3× and 1×), with dashed lines indicated how these oligos are used to assemble a pseudogene, in this example with three different sequences (green, magenta and blue) encoded in a 3:3:1 stoichiometric ratio. The exact order of the sequences is not important and can be shuffled, as in this example, to facilitate gene synthesis. (**b**) Illustration of *in vitro* RCA (left) or helper phage rescue (right) approaches. In the *in vitro* protocol, the pseudogene is cut out with restriction enzymes, re-ligated into smaller circles that are subsequently nicked to form a template for RCA using a Phi29 polymerase. In the phagemid protocol the entire vector, including the pseudogene, is packaged in phage particles in single-stranded form and extracted. In both cases, the final digestion product contains the ODNs derived from a single bacterial colony at the desired stoichiometry. (**c**) Schematic of the hairpin used to produce MOSIC ODNs. A BseGI restriction site (outlined by vertical lines; locations of cuts are marked with blue arrowheads) is flanked on one side by a GAA loop and a GCGC stem, and on the cutting side by two hybridized base pairs, where N is any nucleotide. These base pairs are formed by encoding two complementary bases to the last bases at the 3′ end of ODN 1 into the hairpin sequence. Note that these two bases are encoded in the hairpin byproduct, and consequently that the sequence of the resulting ODNs do not depend on the neighboring ODNs.

RESULTS

The MOSIC method

In our MOSIC method, the starting point is a list of ODN sequences required for the experiment. A Python script (**Supplementary Note**) furnishes a computer-generated DNA construct, which contains all the desired sequences on a line, separated by hairpin regions encoding recognition sequences for a restriction enzyme, in this case, BseGI. If a specific molar ratio of ODNs with respect to each other is required, this information is encoded (or 'hard-coded') in the construct by inserting the required number of sequences for each ODN. For example, ODNs A, B and C could be produced in a 3:3:1 stoichiometric ratio by encoding sequence A three times, sequence B three times and sequence C once (**Fig. 1a**). This computer-generated sequence, the 'MOSIC pseudogene', is then produced by sequence-verified gene synthesis. In gene synthesis, assembly PCR[14] is used to form a long double-stranded construct followed by cloning into *E. coli*, where the construct replicates as plasmid DNA that is subsequently sequenced. The initial construct usually contains errors, but this can be corrected in iterations of sequencing and mutagenesis until the desired correct sequence has been obtained as plasmid DNA derived from a single bacterial colony. For the work presented here, we acquired most of the constructs from commercial providers of gene synthesis services (Online Methods).

After plasmid DNA containing the MOSIC pseudogene has been obtained, the DNA can be amplified in single-stranded form *in vitro* by excision of the pseudogene followed by recircularization and nicking to create templates for RCA (**Fig. 1b**). Alternatively, the DNA can be amplified by cloning into a phagemid vector and subsequent production of single-stranded DNA (ssDNA) via rescue by helper phage (**Fig. 1b**). Both processes (**Supplementary Fig. 1**) furnish ssDNA, digestion of which using a type IIs restriction enzyme releases the ODNs, here referred to as 'MOSIC ODNs'. Using the restriction enzymes BtsCI or BseGI and the hairpin architecture depicted in **Figure 1c**, sequences of consecutive ODNs can be designed to be independent of each other as no base-pairing between the MOSIC ODNs is required for the digestion.

Application to formation of DNA crystal

As an initial proof of concept for the MOSIC method, we encoded the oligonucleotides used to form a rationally designed tensegrity triangle DNA crystal that had been recently published[15]. In that work[15], a triangular structure comprising the 'unit cell' had been designed to fold from three different sequences in a ratio of 3:3:1, using seven individual oligonucleotides. The oligonucleotides form three helices, tailed with short, single-stranded cohesive segments that lead to polymerization and ultimately

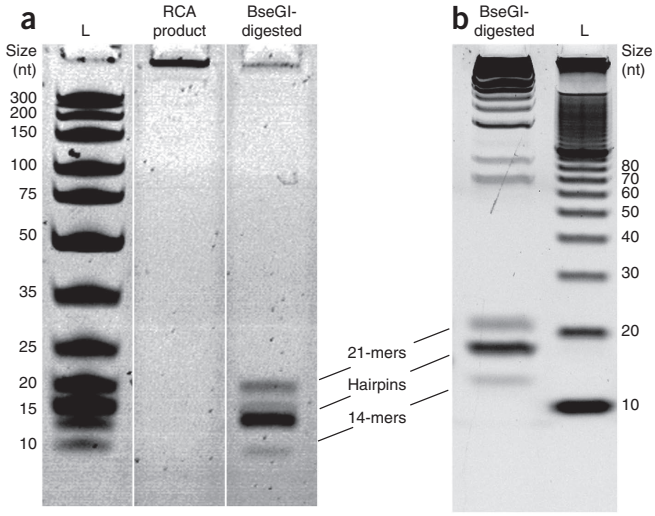


Figure 2 | Digestion of the amplified crystal pseudogene releases the correct length ODNs. (**a**) RCA product after *in vitro* amplification and the same sample after digestion with BseGI. (**b**) BseGI-digested ssDNA amplified in *E. coli* in a phagemid vector. The pseudogene has not been separated from the vector before digestion (high-molecular weight bands are attributed to vector DNA). L, 10-bp ladder; nominal sizes in nucleotides (nt) are indicated. Denaturing polyacrylamide gels were stained with Sybr Gold (uncropped gel for **a** is shown in **Supplementary Fig. 3**).

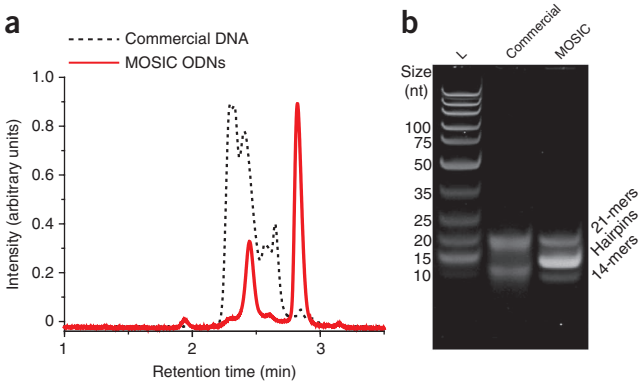


Figure 3 | Comparison of synthetic, commercial DNA with MOSIC ODNs for crystallization. (**a**) HPLC analysis of synthetic, HPLC-purified DNA as provided by a commercial vendor (mixture of the 21-mers and 14-mers (no hairpins included) in the proper relative stoichiometry) and MOSIC ODNs. Two peaks corresponding to the 14-mers (left peak) and 21-mers (right peak) are visible, but double-stranded hairpin sequences exhibited very weak retention in the column and are not visible here. (**b**) Denaturing PAGE of HPLC-purified, synthetic oligonucleotides as provided by a large commercial vendor and MOSIC ODNs. L, ladder; nominal sizes in number of nucleotides (nt) are shown.

a three-dimensional DNA crystal. We chose this model to test the MOSIC method because crystallization generally requires very pure ODNs in a precisely controlled relative stoichiometry. A commercial gene synthesis service provided the pseudogene we designed (similar to that depicted in **Fig. 1a**), containing the required oligonucleotides separated by hairpins (here referred to as the crystal pseudogene; see **Supplementary Table 1** for sequence).

We amplified ssDNA from the clonal template in two ways. In the first approach, we amplified the pesudogene *in vitro* using RCA, where we produced double-stranded plasmid DNA in bacterial culture and purified it using a plasmid miniprep kit. Subsequently, we performed four different enzymatic reactions for *in vitro* amplification: (i) linearization of the crystal pseudogene from the vector by using a nonpalindromic restriction enzyme that recognizes a sequence at the ends of the pseudogene and provides sticky ends, (ii) ligation to circularize the crystal pseudogene, (iii) cleavage by a nicking endonuclease that cuts one of the strands providing a free 3′ OH end, (iv) RCA by using the processive and strand-displacing phi29 DNA polymerase. In the second approach, we grew *E. coli* and extracted helper phage particles. In this case, large amounts of ssDNA can be produced with little effort, but vector DNA is amplified along with the pseudogene (**Supplementary Fig. 2**). In all cases, we digested the resulting ssDNA using BseGI (**Fig. 2**). We also tested BtsCI with similar results.

When run on a polyacrylamide gel, the long ssDNA product of RCA, comprising the crystal pseudogene repeated in tandem remained in the well (**Fig. 2a** and **Supplementary Fig. 3**). The digested products comprised ODNs of the expected length and hairpin by-products (**Fig. 2a**). By measuring the intensity of the gel bands, we verified the 3:3:1 stoichiometric ratio of the ODNs (**Supplementary Fig. 4**). Moreover, we detected no undigested product. The digested phagemid ssDNA (**Fig. 2b**) had product bands of the same size as the *in vitro*–produced MOSIC ODNs, but large amounts of higher-molecular-weight bands were also visible as is to be expected because of the presence of the entire vector ssDNA. In this case we also verified the desired 3:3:1 stoichiometry (**Supplementary Fig. 5**).

For the crystallization experiment, we were interested to determine whether crystals would form without any further purification, and we therefore focused on the *in vitro*–produced variant. We desalted the unpurified, *in vitro*–produced MOSIC ODNs for the tensegrity crystal on a Sep Pak C18 column and subsequently lyophilized them. We compared the quality of these ODNs to that of HPLC-purified synthetic oligonucleotides from a large commercial vendor by running an analytical HPLC analysis (**Fig. 3a** and **Supplementary Fig. 6**) and a denaturing PAGE (**Fig. 3b**). The HPLC peaks of unpurified MOSIC ODNs, and the corresponding bands in a polyacrylamide gel, were sharper than those of control commercial HPLC-purified synthetic oligonucleotides of the same sequence. The apparent polydispersity of the commercial HPLC-purified ODNs was in accordance with previous reports[5].

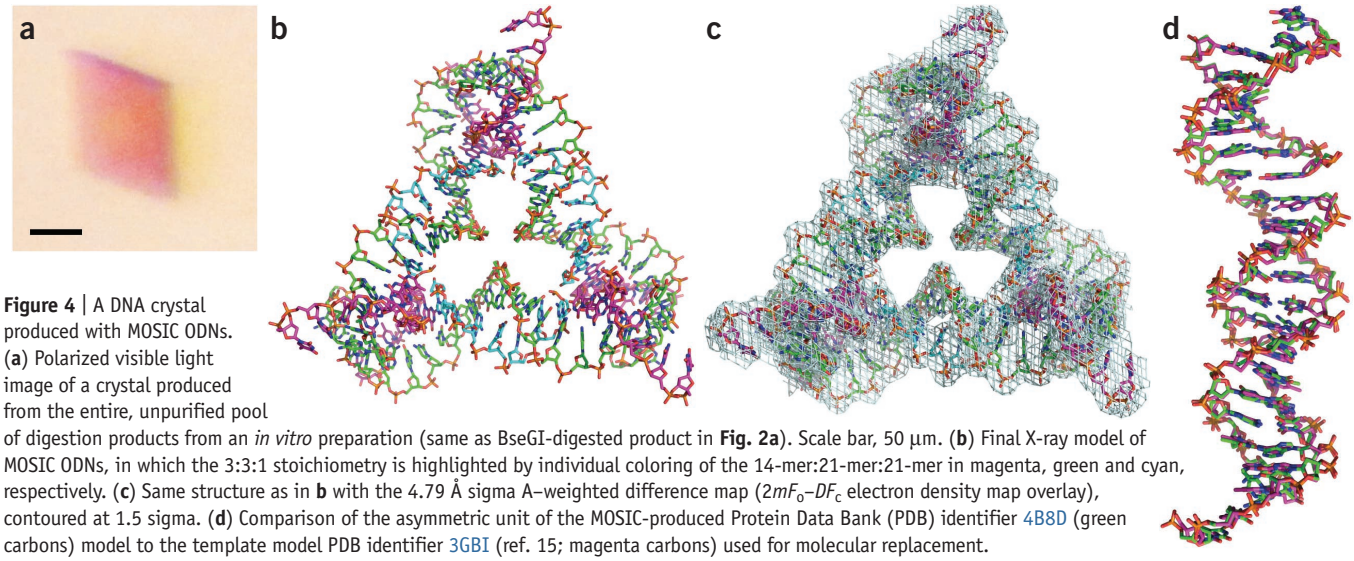


Figure 4 | A DNA crystal produced with MOSIC ODNs. (**a**) Polarized visible light image of a crystal produced from the entire, unpurified pool of digestion products from an *in vitro* preparation (same as BseGI-digested product in **Fig. 2a**). Scale bar, 50 µm. (**b**) Final X-ray model of MOSIC ODNs, in which the 3:3:1 stoichiometry is highlighted by individual coloring of the 14-mer:21-mer:21-mer in magenta, green and cyan, respectively. (**c**) Same structure as in **b** with the 4.79 Å sigma A–weighted difference map ($2mF_o$–DF_c electron density map overlay), contoured at 1.5 sigma. (**d**) Comparison of the asymmetric unit of the MOSIC-produced Protein Data Bank (PDB) identifier 4B8D (green carbons) model to the template model PDB identifier 3GBI (ref. 15; magenta carbons) used for molecular replacement.

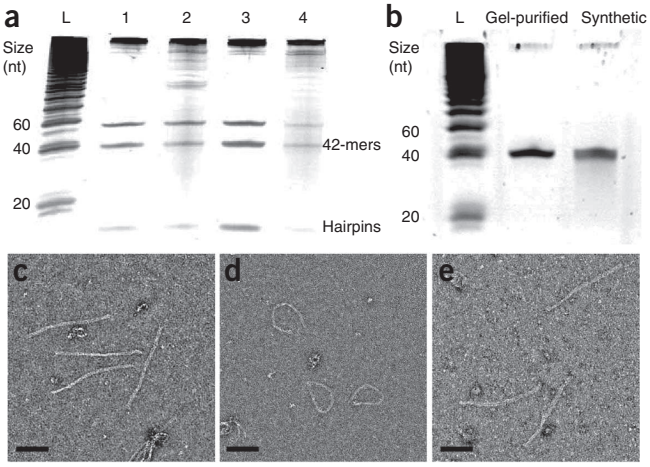
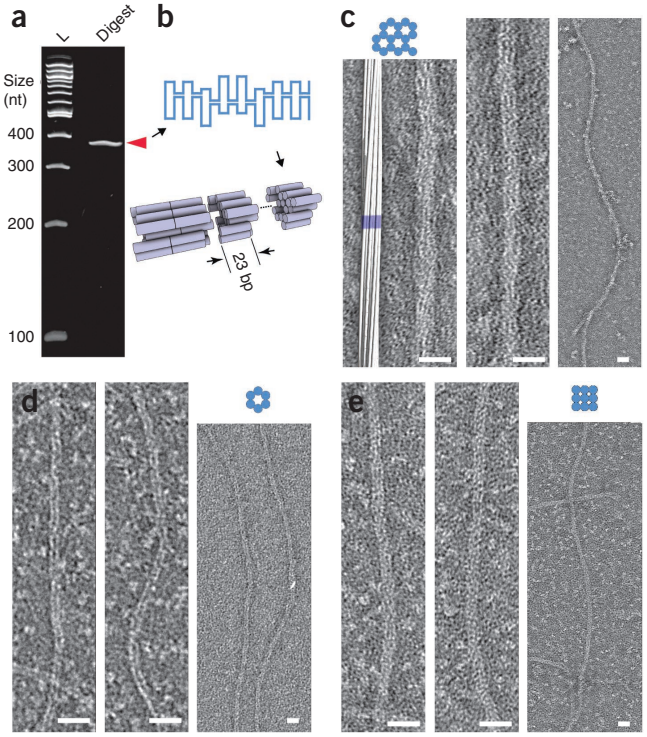


Figure 5 | DNA origami using MOSIC ODNs. (**a**) Denaturing polyacrylamide gel (20%) of digests of 10-helix bundle staple sets 1–4 with BseGI after amplification by helper phage rescue. (**b**) Gel-purified staple set and commercially synthesized staple set 1. L, 20-bp ladder (nominal sizes in nucleotides (nt) indicated). (**c**) TEM image of the 10-helix bundle folded with all synthetic staple strands. (**d**) TEM image of the 10-helix bundle lacking 72 staple strands. (**e**) TEM images of the 10-helix bundle after addition of the lacking 72 staple strands produced by the MOSIC method. Scale bars, 100 nm.

Using a protocol similar to the one described in ref. 15, we prepared samples for crystallization. After 1 week of cooling from 60 °C to room temperature in a sitting drop plate, we observed rhombohedron-shaped crystals with dimensions of ~70–150 µm (**Fig. 4** and **Supplementary Fig. 7**). We acquired X-ray diffraction data and found that the structures determined by molecular replacement closely matched (**Fig. 4d**) the previously published structures of DNA crystals[12,15]. Similarly to results in ref. 12, because of the enzymatic production method used, our data showed 5′-end phosphates (**Supplementary Fig. 8**) lacking in the original published structure derived from synthetic ODNs[15]. The resolution obtained in our study using conventional beamlines is slightly lower compared to that reported in refs. 12 and 15. The 5′-end phosphates have been shown to improve the resolution of these types of crystals[12]. Our findings do not contradict this, rather, our data are not comprehensive enough to make conclusions about whether 5′-end phosphates improve resolution or not.

Application to DNA nanotechnology

As an additional proof of concept, we produced MOSIC ODNs for the folding of DNA origami nanostructures. In this case we used the phagemid production strategy (**Fig. 1b**).



One of the advantages of the MOSIC phagemid method is the possibility to scale up the production of many high-quality ODNs. The development of technologies based on self-assembled DNA nanostructures and in particular the introduction of DNA origami method involves the use of many oligonucleotides[16–19]. Certain applications, such as the production of alignment media for the determination of protein structure[20] or the folding of complex DNA nanostructures used for biological applications[21–23] require many oligonucleotides of the same sequence. With these considerations in mind, we therefore selected the folding of a DNA origami structure to test the capabilities of phagemid-produced MOSIC ODNs.

We designed a honeycomb 10-helix bundle using 180 staple ODNs, each 42 nt long. Seventy-two of the ODNs were encoded on four different DNA pseudogene constructs as described above. Each pseudogene contained 18 staples, here referred to as a 'staple set'. We cloned the staple sets into phagemid vectors (pBluescriptIISK(−) or pGEM-Teasy), followed by amplification in *E. coli* and rescue via helper phage (**Supplementary Fig. 9**). We then digested the resulting ssDNA with BseGI (**Fig. 5a**) and purified the DNA by PAGE to remove byproducts from the vector (**Fig. 5b**). We used the obtained staple strands in DNA origami folding reactions.

The 72 strands we selected for production by the MOSIC method lie along the length of the tube structure on one side (**Supplementary Figs. 10** and **11**). Because of this design, structures folded without these 72 ODNs should lack rigidity on

Figure 6 | Applying the MOSIC method for production of very long DNA oligonucleotides. (**a**) Denaturing 5% polyacrylamide gel showing the digestion product of a pseudogene that releases a 378 nt MOSIC ODN (arrowhead). L, 100-bp ladder, with the nucleotide (nt) count indicated. (**b**) Schematic for ultra-small origami folding: 378-nt strands were used as scaffold and folded into 17 helices forming a 23-base-pair-long, ultra-small DNA origami brick that polymerizes into an extended 17-helix bundle. (**c**) TEM micrographs of the polymerized ultra-small DNA origami, and a CanDo structure prediction[31] (top; to scale) of 25 monomers. One monomer is highlighted in blue. The two left micrographs show close-ups of two different tubes where the internal structure is visible. The rightmost micrograph shows 850 nm of tube corresponding to over 100 polymerized monomers. (**d**,**e**) Data as in **c** for similar designs, where the 378-nt MOSIC ODN was folded into 6-helix bundles or 9-helix bundles and polymerized. Scale bars, 20 nm.

one side. Indeed, by transmission electron microscopy (TEM) imaging we observed formation of structures that appeared to curl up when we excluded these 72 staple ODNs (**Fig. 5c–e**). In contrast, when we added the missing strands, ODNs produced by MOSIC, the straight original design (**Fig. 5c**) reappeared (**Fig. 5e** and **Supplementary Fig. 12**), indicating that the staple sets produced by the MOSIC method can replace synthetically produced oligonucleotides in DNA origami applications.

Application to production of long ODNs

It is challenging to produce long ODNs comprising several hundred nucleotides via current synthetic methods. Using the MOSIC method, however, very long oligonucleotides can be readily extracted by simply increasing the distance between the cutter hairpins in the clonal substrate. Using the MOSIC method, we encoded a 378-nt strand on a clonal template flanked by the same cutter hairpins as described above. We extracted large amounts of 378-nt ODNs (**Fig. 6a** and **Supplementary Fig. 13**). Compared to the few available sources of synthetic DNA of that length, the MOSIC method yields single-stranded ODNs at a considerably lower cost (**Supplementary Fig. 13**).

We tested the functionality of long ODNs produced by the MOSIC method using the 378-nt strands, without additional purification, as a scaffold for folding ultra-small DNA-origami structures (**Fig. 6b–e** and **Supplementary Figs. 14** and **15**). We tested three constructs of varying width, in which the 378-nt ODN folded into short 17-, 9- or 6-helix bundles. Each ultra-small DNA origami is expected to form a small brick, ~10 nm × 9 nm × 8 nm for the 17-helix bundle. To facilitate TEM imaging, we designed the monomeric structures to polymerize head to tail using protruding staple sticky ends. TEM imaging of the folded structures revealed long polymeric nanotubes (**Fig. 6c–e** and **Supplementary Fig. 16**). We implemented the 17-helix bundle design with a DNA twist of 11 base pairs per turn, previously reported to increase polymerization efficiency[24], leading to a global right-handed twist of the polymers. The structures had the expected width and internal structure.

DISCUSSION

ODNs are an important molecular tool for many important research areas: as probes for cell biology and diagnostics, as primers for molecular genetics or as components for DNA nanotechnology and DNA computing. ODNs are also being used to develop therapeutics, with one ODN-based drug already approved[25,26]. Given the demand for high quality ssDNA oligomers, the method we reported solves several problems associated with synthetic ODNs.

MOSIC ODNs originate from sequence-verified templates from a single colony of bacteria and are monodisperse. The correct folding of a DNA nanostructure, and in particular, the crystal structure data, are clear evidence of the correct sequence in the MOSIC ODNs. In contrast to previous work[15], where two consecutive HPLC purification steps were used to prepare synthetic ODNs, we observed crystallization directly of our pseudogene digest without any purification except desalting and enzyme removal. We estimated per-base error rates in MOSIC ODNs to be equivalent to the replication error rates for phi29 polymerase ($\sim 10^{-5}$–10^{-6})[27] or to the error rates of the bacterial replication machinery ($\sim 10^{-7}$–10^{-8})[28] for the *in vitro* and phagemid-based protocols, respectively.

MOSIC ODNs can be produced in a precisely controlled relative stoichiometry, which is easily implemented by encoding the ratios in the pseudogenes. The crystallization, directly after a digest of a pseudogene, where an exact 3:3:1 ratio is necessary for assembly of the structure[15], is indicative of the capabilities of the MOSIC method in this regard. The stoichiometry control could prove to be important for producing small therapeutic DNA structures, such as a recently proposed method for silencing of a tumorigenic gene[29].

The MOSIC method provides the flexibility of choosing whether to amplify *in vitro* or in *E. coli*. Although we did not investigate this, it is reasonable to assume that the phagemid-based procedure would result in methylated ODNs when non-methylation–deficient bacterial strains are used. Depending on the application, this could be useful or viewed as a disadvantage. The *in vitro* procedure gives a completely clean digest but is slightly more labor-intensive and requires the purchase of more enzymes than the phagemid procedure.

The MOSIC method can be applied to produce ODNs comprising several hundred bases (we produced 14–378-nt ODNs). This allowed us to make small, scaffolded origami constructs that to our knowledge have not been demonstrated before. The MOSIC method therefore opens a path to construct DNA nanostructures using ODNs much longer than the standard oligonucleotides (<40–60 nt long) normally used for these types of applications. MOSIC could be used to produce designs that are a hybrid between the recently reported single-stranded tile[30] and scaffolded-origami[16] methods.

Finally, long MOSIC ODNs can be produced at a much lower cost than commercially synthesized ODNs of the same length, about 15–30 times cheaper for a 378-nt ODN (excluding initial cost for gene synthesis of templates; **Supplementary Fig. 13**). Our calculations are based on listed prices for small-scale amounts of all involved enzymes.

In contrast to enzymatic production proposed before[9–11,13], the MOSIC method does not rely on the addition of any synthetic primers, neither for amplification nor for digestion. In fact, regardless of whether amplification is done *in vitro* or *in vivo* by *E. coli* and phage, all the components necessary for mass-production of MOSIC ODNs can be readily grown in bacterial cultures. Taken together, the MOSIC method provides a route for cheap, abundant production of oligonucleotides for any application that requires high-quality ODNs derived from a single bacterial colony and in a defined stoichiometry.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Protein Data Bank: 4B8D (tensegrity crystal from MOSIC ODNs).

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

Funded by the Swedish Research Council (Vetenskapsrådet) through a repatriation grant and a project grant to B.H. (grants 2010-6296 and 2010-5060). B.H. is a recipient of an assistant professorship with startup funding by Carl Bennet AB, Karolinska Institutet and Vinnova. We thank S. Douglas for help with the 10-helix bundle design, members of the Laboratory of Chemical Biology at Karolinska Institutet for HPLC support, T. Karlberg and members of the Protein Science Facility for crystallography support, M. Schultz and C. Sandén for help

with pseudogene cloning and enzymatic reactions for the long ODNs, and N. Seeman for fruitful discussions.

AUTHOR CONTRIBUTIONS

C.D., C.K. and B.H. contributed to experiments. M.M. contributed to the crystallography experiments. B.H. conceived the method principle, and B.H. and W.M.S. contributed to the method design. C.D. and C.K. contributed to method development and implementation. All authors contributed to figure production and manuscript writing.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Merrifield, R.B. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).
2. The International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
3. Badi, N. & Lutz, J.F. Sequence control in polymer synthesis. *Chem. Soc. Rev.* **38**, 3383–3390 (2009).
4. Oberacher, H., Niederstätter, H. & Parson, W. Characterization of synthetic nucleic acids by electrospray ionization quadrupole time-of-flight mass spectrometry. *J. Mass Spectrom.* **40**, 932–945 (2005).
5. Semenyuk, A. *et al.* Cartridge-based high-throughput purification of oligonucleotides for reliable oligonucleotide arrays. *Anal. Biochem.* **356**, 132–141 (2006).
6. Zhang, D.Y., Turberfield, A.J., Yurke, B. & Winfree, E. Engineering entropy-driven reactions and networks catalyzed by DNA. *Science* **318**, 1121–1125 (2007).
7. Yin, P., Choi, H.M.T., Calvert, C.R. & Pierce, N.A. Programming biomolecular self-assembly pathways. *Nature* **451**, 318–322 (2008).
8. Zhang, D.Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nat. Chem.* **3**, 103–113 (2011).
9. Tian, J. *et al.* Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).
10. Antson, D.O., Isaksson, A., Landegren, U. & Nilsson, M. PCR-generated padlock probes detect single nucleotide variation in genomic DNA. *Nucleic Acids Res.* **28**, E58 (2000).
11. Kosuri, S. *et al.* Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol.* **28**, 1295–1299 (2010).
12. Sha, R. *et al.* Self-assembled DNA crystals: the impact on resolution of 5′-phosphates and the DNA source. *Nano Lett.* **13**, 793–797 (2013).
13. Lohmann, J.S., Stougaard, M. & Koch, J. A new enzymatic route for production of long 5′-phosphorylated oligonucleotides using suicide cassettes and rolling circle DNA synthesis. *BMC Biotechnol.* **7**, 49 (2007).
14. Stemmer, W.P.C., Crameri, A., Ha, K.D., Brennan, T.M. & Heyneker, H.L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).
15. Zheng, J. *et al.* From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature* **461**, 74–77 (2009).
16. Rothemund, P.W.K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
17. Douglas, S.M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
18. Högberg, B., Liedl, T. & Shih, W.M. Folding DNA origami from a double-stranded source of scaffold. *J. Am. Chem. Soc.* **131**, 9154–9155 (2009).
19. Andersen, E.S. *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459**, 73–76 (2009).
20. Douglas, S.M., Chou, J.J. & Shih, W.M. DNA-nanotube-induced alignment of membrane proteins for NMR structure determination. *Proc. Natl. Acad. Sci. USA* **104**, 6644–6648 (2007).
21. Douglas, S.M., Bachelet, I. & Church, G.M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831–834 (2012).
22. Zhao, Y.-X. *et al.* DNA origami delivery system for cancer therapy with tunable release properties. *ACS Nano* **6**, 8684–8691 (2012).
23. Jiang, Q. *et al.* DNA origami as a carrier for circumvention of drug resistance. *J. Am. Chem. Soc.* **134**, 13396–13403 (2012).
24. Dietz, H., Douglas, S. & Shih, W.M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325**, 725 (2009).
25. Patil, S.D., Rhodes, D.G. & Burgess, D.J. DNA-based therapeutics and DNA delivery systems: a comprehensive review. *AAPS J.* **7**, E61–E77 (2005).
26. Rosi, N.L. *et al.* Oligonucleotide-modified gold nanoparticles for intracellular gene regulation. *Science* **312**, 1027–1030 (2006).
27. Esteban, J.A., Salas, M. & Blanco, L. Fidelity of phi29 DNA Polymerase. *J. Biol. Chem.* **268**, 2719–2726 (1993).
28. Schaaper, R.M. Base selection, proofreading, and mismatch repair during DNA replication in *Escherichia coli*. *J. Biol. Chem.* **268**, 23762–23765 (1993).
29. Lee, H. *et al.* Molecularly self-assembled nucleic acid nanoparticles for targeted *in vivo* siRNA delivery. *Nat. Nanotechnol.* **7**, 389–393 (2012).
30. Wei, B., Dai, M. & Yin, P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **485**, 623–626 (2012).
31. Castro, C.E. *et al.* A primer to scaffolded DNA origami. *Nat. Methods* **8**, 221–229 (2011).

ONLINE METHODS

Pseudogene designs. We ordered the crystal pseudogene (CP) from a gene-synthesis provider (MrGene), which provided it directly as a double-stranded plasmid construct in a pMK-RQ vector. The CP is enclosed by two restriction sites, CGTCTC, recognized by the restriction enzyme BsmBI, and the circularized form is characterized by a nicking site, GCAATG, recognized by Nb.BsrDI (**Supplementary Fig. 17a**). The seven oligonucleotides are separated by eight hairpins characterized by a 3-base loop, a 4-base GCGC stem, the BseGI restriction site GGATG and two variable overhanging bases that are complementary to the last two bases of the preceding encoded oligonucleotide (**Fig. 1d**).

We made the 378-nt-long ODN pseudogene in our laboratory by performing a PCR with pBluscriptSKII(+) as template and two partially homologous primers (forward: 5′-GGTCTCACATTGC ATAATTAACATCCGCGGAACGCGGATGTTCCGGCGTCAA TACGGGATAA-3′; reverse, 5′-GGTCTCCAATGCAGAAGAGC TGCATCCGCGTTCCGCGGATGCACTTTTCGGGGAAATGT GC-3′). To perform the *in vitro* amplification, the pseudogene was enclosed by two restriction sites, GGTCTC, which is recognized by BsaI, and the circularized form is characterized by two nicking sites, GCAATG and GCTCTTC, recognized respectively by Nb.BsrDI and Nt.BspQI, which cut at 1-base distance on the same strand (**Supplementary Fig. 17b**). The 378-nt-long pseudogene is also characterized by two hairpins like those used for the CP.

We designed the 10-helix bundle nanotube using the software caDNAno[32] in accordance with the design rules for three-dimensional DNA origami on a honeycomb lattice reported previously[17]. The staple strands were sorted into pools of 18 strands resulting in 10 staple sets (**Supplementary Fig. 10**). Staple sets 1–4 were assembled by gene synthesis according to a similar scheme used for the crystal ODNs as described above. The staple-set pseudogenes were obtained directly from commercial gene-synthesis services as a double-stranded plasmid constructs in a pBluescript SKII(−) vector (Genewiz, staple sets 1 and 2) or a pGEM-T easy vector (Bioneer, staple sets 3 and 4). The sequences corresponding to the staple-strand ODNs were separated by 19 hairpin sequences characterized by a 3-base loop (GAA), a 3-base stem (CGC), the BseGI restriction site GGATG and two variable overhanging bases complementary to the last two bases of the preceding encoded oligonucleotide. The pseudogenes for staple sets 1 and 2 were enclosed by the restriction sites G′GATCC, recognized by the restriction enzyme BamHI and T′CTAGA, recognized by the restriction enzyme XbaI, where ′ indicates the site of digest.

Transformation of *E. coli* cells with a vector containing the pseudogenes. We inserted the linear CP DNA into a pBluescript SK II(−) for the phage cloning strategy, and we used the pMK-RQ containing the CP for the *in vitro* amplification. We cloned the 378-nt-long ODN pseudogene into a pDrive cloning vector (Qiagen, PCR cloning kit). We thawed SCS110 or XL10 gold competent cells (Agilent technologies) on ice. For each reaction, we gently mixed and incubated 100 µl of cells with 1.7 µl of β-mercaptoethanol (1.42 M) for 10 min, swirling gently every 2 min. Then we mixed 50 ng of the vector containing the pseudogene, or pUC18 plasmid as a control, with competent cells and incubated for 30 min on ice. Then we heat-pulsed the reactions in a 42 °C water bath for 45 s and incubated on ice again for 2 min. We then added 0.9 ml of preheated (42 °C) NZY$^+$ broth (10 g of NZ amine (casein hydrolysate), 5 g of yeast extract, 5 g of NaCl in 1 l of water subsequently adjusted to pH 7.5 using NaOH, supplemented with 12.5 ml of 1 M $MgCl_2$, 12.5 ml of 1 M $MgSO_4$ and 20 ml of 20% (w/v) glucose), and we incubated the reactions at 37 °C for 1 h at 250 r.p.m.

We plated 100 µl of the transformation mixture on LB agar plates containing kanamycin (50 µg/ml) for pMK-RQ–transformed cells and ampicillin (100 µg/ml) for the pBluescript SK II(−), pDrive cloning vector and pUC18 plasmid. We incubated the plates at 37 °C overnight. The day after, we screened single colonies, from pMK-RQ and pBluescript SK II(−)–transformed cells, for the CP insert by BsmBI digestion, and by BsaI digestion for the 378-nt-long ODN pseudogene. Complete sequence of the 378-nt pseudogene, verified by sequencing, is available in **Supplementary Table 2**.

We transformed *E. coli* (SCS110) with staple pseudogenes as described above, plating the transformed cells on LB agar plates containing ampicillin (100 µg/ml) and screening single colonies DNA for the pseudogenes. We used XL10 gold cells in all cases except when producing phagemid DNA; then we used SCS110 cells.

Pseudogene *in vitro* amplification. We inoculated 5 ml of LB containing kanamycin (50 µg/ml) with a single colony of *E. coli* transformed with the pMK-RQ vector containing the CP pseudogene and incubated overnight at 37 °C, shaking at 250 r.p.m. We isolated the plasmid DNA from the saturated cell culture by a plasmid 'mini-prep' kit (Omega Bio-Tek), and digested the pMK-RQ vector containing the CP (20 ng/µl) by BsmBI (0.25 U/µl) (New England Biolabs), in 1× NEB2 buffer reaction at 55 °C for 2 h, followed by heat inactivation at 80 °C for 20 min. We loaded the digestion products on a 1% agarose gel containing ethidium bromide (1 µg/ml; Sigma Aldrich), and purified the linear pseudogene (329 base pairs) by gel extraction (kit from Omega bio-tek) and eluted in 30 µl of 10 mM Tris-HCl (pH 8.5). We applied the same procedure to get the linear 378-nt-long ODN pseudogene but used ampicillin (100 µg/ml) in the 5 ml LB culture and BsaI as restriction enzyme (New England Biolabs). We ligated the linear pseudogene (5 ng/µl) by T4 ligase (0.25 U/µl; Fermentas) in 1× rapid ligation buffer at 22 °C for 10 min, followed by an inactivation step at 65 °C for 10 min. We nicked the resulting circular pseudogene (1 ng/µl) by Nb.BsrDI (0.5 U/µl, New England Biolabs) (with addition of Nt.BstQI at the same concentration for the circular 378-nt-long ODN pseudogene) in 1× NEB2 buffer reaction at 65 °C for 2 h; we stopped the reaction by heating at 80 °C for 25 min. We amplified the nicked circular pseudogene (0.25 ng/µl) overnight at 30 °C by rolling circle amplification (RCA) in a single-stranded form by phi29 DNA polymerase (0.5 U/µl; Fermentas) in a 1× phi29 reaction buffer reaction (33 mM Tris acetate, 10 mM magnesium acetate, 66 mM potassium acetate, 0.1% (v/v) Tween 20 and 1 mM DTT; Fermentas) containing dNTPs mix (1 mM each; Fermentas) and T4 gene 32 ssDNA-binding protein (0.1 µg/µl, New England Biolabs). All these reactions (**Supplementary Fig. 1**) do not require any desalting steps; instead we diluted the solutions from one step to the other, adding the new buffer reaction.

In a typical experiment we ligated 100 ng of linear pseudogene in 20 µl, performed the nicking reaction in 100 µl and performed the RCA in 400 µl. In the final BseGI digestion step, we diluted the

slimy RCA solution into 1.6 ml, producing MOSIC ODNs typically around 12 ng/µl (see below for BseGI digestion details).

Pseudogene amplification in bacteria. We grew phagemid-transformed single colonies in 25 ml LB agar with ampicillin (100 µg/ml) at 37 °C overnight. We then inoculated four 300-ml cultures of 2× YT (16.0 g/l tryptone, 10.0 g/l yeast extract and 5.0 g/l NaCl, adjusted to pH 7.0 with NaOH) medium (containing 100 µg/ml ampicillin) with 3 ml of the saturated overnight culture, and added 1.5 ml of sterile 1 M $MgCl_2$. We incubated the culture at 37 °C with shaking at 250 r.p.m. and monitored the OD_{600} during the cell growth. We added VCSM13 helper phage at OD_{600} 0.45 corresponding to around 2.5×10^8 cells/ml with a multiplicity of infection of 20 (ratio of phage to cells). After an additional 5 h, we discarded the bacteria by centrifuging twice at 3,800*g* for 20 min at 4 °C. Afterward we dissolved PEG 8000 (40 g/l; Sigma-Aldrich) and NaCl (30 g/l; Sigma-Aldrich) in the recovered supernatant and incubated it for 1 h on ice. We spun down the phage in this cloudy solution at 15,000*g* for 30 min, and after discarding the supernatant we resuspended the pellet in 6 ml Tris (10 mM, pH 8.5). We centrifuged the collected suspension at 3,800*g* for 5 min at 4 °C to get rid of any bacterial residue, and we transferred the supernatants to fresh centrifuge bottles. To remove the phage proteins, we added a double volume of (0.2 M NaOH and 1% SDS) to the solution, and after gently swirling, we incubated at room temperature for 3 min. We added a one and a half volume of (3 M KOAc pH 5.5 titrated with glacial acetic acid), gently mixed the solution by inversion and incubated it in ice-cold water for 10 min. After centrifuging at 16,500*g* for 30 min at 4 °C, we decanted the supernatant to a fresh bottle and added at least two volumes of 100% ethanol. After 1 h incubation in ice-cold water, we performed another centrifugation at 16,500*g* for 30 min. The obtained single-stranded DNA (pellet), was washed with 70% ethanol and resuspended in 2 ml Tris (10 mM pH 8.5) and analyzed by agarose gel (**Supplementary Figs. 2** and **9**).

Oligonucleotide production. We digested the RCA product (up to 30 ng/µl) and the DNA from bacterial production (52 ng/µl) by BseGI (0.5 U/µl, Fermentas) in 1× Tango buffer by incubating at 55 °C for 24 h followed by heat-inactivation at 80 °C for 20 min. We analyzed the digestion products by PAGE (20% polyacrylamide, 20% formamide, 8 M urea mixed in 1× TBE which is made of 89 mM Tris-borate, 2 mM Na_2EDTA dissolved in deionized water) at 180 V for 1 h (**Fig. 2a**,**b**) and for the 378-nt ODN pseudogene by PAGE (5% polyacrylamide, 20% formamide, 8 M urea mixed in 1× TBE), at 180 V for 35 min (**Fig. 6a**). We stained the gels with SYBR Gold (1×; Invitrogen) for 15 min and acquired images by UV trans-illumination (UVITEC) and analyzed by the software Image J. We also analyzed the MOSIC CP oligonucleotides by HPLC, and we compared them with the synthetic corresponding oligonucleotides. We performed analytical reverse-phase separations of oligonucleotides on an Agilent 1100 system. We used a C18 column (X bridge, 3.0 mm × 50 mm) from Waters in a system of buffer A (10 mM NH_4HCO_3 in H_2O, pH 10) and buffer B (MeCN) at a flow of 1 ml/min (buffer gradient from 1% to 11% buffer B in 4 min).

Crystallization experiment. We desalted oligonucleotides from the RCA digest, after StrataClean Resin (Agilent) extraction to remove proteins, by Sep-Pak C18 cartridges (Waters) following the manufacturer's instructions. We subsequently lyophilized them, and resuspended them in crystallization buffer. Crystals were grown from 50-µl sitting drops in a thermally controlled incubator containing 250 ng/µl DNA, 30 mM sodium cacodylate, 50 mM magnesium acetate, 50 mM ammonium sulfate, 5 mM magnesium chloride and 25 mM Tris (pH 8.5), equilibrated against a 1 ml reservoir of 1.7 M ammonium sulfate. Rhombohedron-shaped crystals with dimensions between 70 µm and 150 µm were obtained by slow annealing: temperature was decreased from 60 °C to room temperature (~20 °C) with a cooling rate of 0.2 °C per hour over 8 d, during which the volume of the drop diminished by about 90%. Crystals were obtained at the end of the cooling step, and appeared full-sized within a day. We transferred the crystals to a cryosolvent of 30% glycerol, 100 mM ammonium sulfate, 10 mM $MgCl_2$ and 50 mM Tris and froze them by immersion in liquid nitrogen.

Data collection and structure determination. We collected a native data set to 4.79 Å resolution from a single crystal at BESSY BL14-1 in Berlin[33]. (We performed three crystallization droplet experiments, each one gave ample amounts of crystals. In total, we collected three data sets from three crystals one from BESSY BL14-1 and two from Diamond I02, not used in this study.) Our MOSIC DNA oligonucleotides crystallized in space group *H*3 (ref. 34), with cell parameters $a = b = 106.4$ Å and $c = 95.15$ Å (PDB: 4B8D) and the solid-phase–produced oligonucleotides with identical DNA sequence used for the first tensegrity triangle crystal structure[15] (PDB: 3GBI) also crystallized in *H*3 with almost identical cell parameters, $a = b = 107.2$ Å and $c = 93.14$ Å. The 4.79 Å data included 1,980 reflections and the structure was redetermined by molecular replacement using 3GBI as search model and MOLREP integrated in the CCP4i software package[35]. Molecular replacement parameters are given in **Supplementary Table 3**. After initial refinement of the 3GBI model, difference density for the 5′ phosphate of our MOSIC ODNs was visible in our 4.79 Å difference map (**Supplementary Fig. 8**) and the 5′ phosphate was subsequently included in our model being the sole addition made to the 3GBI template model. During refinement in autoBUSTER[36] the crystallographic *B*-factor modeling disorder was kept constant for all atoms of the model as the 4.79 Å resolution data does not permit refinement of individual *B* factors. The resulting DNA tensegrity model contains 866 nucleic acid atoms refined to crystallographic *R* and R_{free} factors of 0.185 and 0.205, respectively. The asymmetric unit was kept the same as during 3GBI refinement, and our structure does not differ substantially from the original 3GBI structure when superimposed (**Fig. 3d**). We generated all the molecular model figures using PyMOL (W.L. deLano, The PyMOL Molecular Graphics System version 1.4.1. Schrödinger LLC, http://www.pymol.org/; 2002).

Folding of 10-helix bundle and electron microscopy. We purified staple sets 1–4 for folding of 10-helix bundle by PAGE (20% polyacrylamide, 20% formamide and 8 M urea). The oligonucleotides were washed out of the gel with water, then desalted by Sep-Pak C18 cartidges (Waters), lyophilized and redissolved in water.

We prepared each sample by combining 5 nM scaffold (p8064), 25 nM of each staple oligonucleotide, buffer and salts including

doi:10.1038/nmeth.2503

5 mM Tris, 1 mM EDTA (pH 7.8 at 20 °C) and 10 mM $MgCl_2$. For the staple strands produced by MOSIC, the concentration was raised to 75 nM. We carried out the folding by rapid heat-denaturation followed by slow cooling (80 °C to 60 °C in 20 min, followed by cooling from 60 °C to 24 °C over 14 h).

We spotted 5 µl of the folding reaction on glow-discharged, carbon-coated Formvar grids (Electron Microscopy Sciences), incubated them for 20 s, blotted off and stained with 2% (w/v) aqueous uranyl formate solution. We performed electron microscopy analysis using a FEI Morgagni 268(D) transmission electron microscope at 80 kV with nominal magnifications between 28,000 and 44,000. We recorded images digitally using the Advanced Microscopy Techniques Image Capture Engine 5.42.

Folding and polymerization of ultra-small DNA origami bricks. After StrataClean Resin (Agilent) extraction, we desalted the 378-nt-long oligonucleotide from RCA-prepared, digested product by Sep-Pak C18 cartridges (Waters) following the manufacturer's instructions. We lyophilized and resuspended the 378-nt ODN in deionized water to be used as scaffold for ultra-small DNA origami bricks. We performed the folding reaction by combining 0.1 µM of the long strand with 0.5 µM staple oligonucleotides (including protruding ones for the polymerization of the origami bricks), buffer and salts including 5 mM Tris, 1 mM EDTA (pH 7.8 at 20 °C) and 10 mM $MgCl_2$. We carried out the folding by rapid heat denaturation followed by slow cooling (80 °C to 60 °C in 20 min, followed by cooling from 60 °C to 24 °C over 14 h).

We spotted 3 µl of the folding reaction on glow-discharged, carbon-coated Formvar grids (Electron Microscopy Sciences), incubated them for 20 s, blotted them off and stained with 2% (w/v) aqueous uranylformate solution. We performed electron microscopy analysis as described above.

32. Douglas, S.M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res.* **37**, 5001–5006 (2009).
33. Mueller, U. *et al.* Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. *J. Synchrotron Radiat.* **19**, 442–449 (2012).
34. Grosse-Kunstleve, R.W., Echols, N. & Adams, P.D. Fuzzy space group symbols: H3 and H32. *Computational Crystallography Newsletter* **2**, 12–14 (2011).
35. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
36. Bricogne, G. *et al. BUSTER version 2.11.2.* (Global Phasing Ltd., 2011).