Self-assembly of genetically encoded DNA-protein hybrid nanoscale shapes

Florian Praetorius and Hendrik Dietz*

INTRODUCTION: Controlling the spatial arrangement of functional components in biological systems on the scale of higher-order macromolecular assemblies is an important goal in synthetic biology. Achieving this goal could yield new research tools and pave the way for interesting applications in health and biotechnology. DNA origami enables constructing arbitrary shapes on the desired scale by folding a single-stranded DNA “scaffold” into user-defined shapes using a set of “staple” oligonucleotides, but single-stranded DNA is typically not available in living cells, and the structures usually assemble only at nonphysiological temperatures. The scope of functions that can be fulfilled by DNA itself appears so far still limited. Proteins, on the other hand, offer a large variety of functionalities and are easily accessible in cells through genetic encoding, but designing larger structural frameworks from proteins alone remains challenging.

RATIONALE: Here, we reimagined DNA origami to explore the possibility for using a set of designed proteins to fold a double-stranded DNA template into user-defined DNA-protein hybrid objects with dimensions on the desired 10- to 100-nm scale. To realize our idea, we require synthetic DNA “looping” proteins that can link two user-defined double-helical DNA sequences, and we need to determine suitable rules for arranging both the template and multiple of such staple proteins in the context of a larger target structure. Transcription activator-like (TAL) effector proteins are produced by plant pathogenic bacteria and injected into host cells, where they bind to specific promoter regions, thus controlling the expression of target genes. The DNA recognizing part of a TAL effector consists of an array of repeat subunits that binds to the major groove of double-helical DNA, thus forming a superhelix. Each repeat subunit comprises ~34 amino acids and recognizes a single DNA base pair. Because of their modular architecture, TAL effectors can be engineered to bind to user-defined DNA sequences, and this technique is currently being exploited for genome engineering applications. For constructing the staple proteins, we therefore chose the DNA-recognizing domains of TAL effectors.

RESULTS: We characterized the DNA-recognizing domains of the TAL effectors with respect to binding affinity and sequence specificity. To construct the staple proteins, we fused two TAL proteins via a custom peptide linker and tested for the ability to connect two separate double-helical DNA domains. For creating larger objects containing multiple staple protein connections, we identified a set of rules regarding the optimal spacing between these connections. On the basis of these rules, we could create megadalton-scale objects that realize a variety of structural motifs, such as custom curvatures, vertices, and corners. Each of those objects was built from a set of 12 double-TAL staple proteins and a template DNA double strand with designed sequence. We also tested design principles for multilayer structures with enhanced rigidity. All components of our nanostuctures can be genetically encoded and self-assemble isothermally at room temperature in near-physiological buffer conditions. The staple proteins used in this work also carried a green fluorescent protein domain that serves as a placeholder for a variety of functional protein domains that can be genetically fused to the staple proteins. We were also able to demonstrate formation of our structures starting from genetic expression in a one-pot reaction mixture that contained the double-stranded DNA scaffold, the genes encoding the staple proteins, RNA polymerase, ribosomes, and cofactors for transcription and translation. Successful self-assembly of our hybrid nanostructures was confirmed using transmission electron microscopy.

CONCLUSION: By using our system of designing double-TAL staple proteins that fold a template DNA double strand, researchers can control the spatial arrangement of protein domains in custom geometries. Our system should have a good chance to work inside cells, given the success of our in vitro expression experiments and considering that the DNA binding properties of TAL are preserved inside cells, as seen in gene editing experiments with TAL-based endonucleases. Because TAL-based staple proteins can be tailored to specifically recognize any desired DNA target sequence, these proteins could then be used to create custom structures and loops in genomic DNA to study the relation between genome architecture and gene expression, or to position proteins involved in other intracellular processes in user-defined ways.
We describe an approach to bottom-up fabrication that allows integration of the functional diversity of proteins into designed three-dimensional structural frameworks. A set of custom staple proteins based on transcription activator–like effector proteins folds a double-stranded DNA template into a user-defined shape. Each staple protein is designed to recognize and closely link two distinct double-helical DNA sequences at separate positions on the template. We present design rules for constructing megadalton-scale DNA-protein hybrid shapes; introduce various structural motifs, such as custom curvature, corners, and vertices; and describe principles for creating multilayer DNA-protein objects with enhanced rigidity. We demonstrate self-assembly of our hybrid nanostructures in one-pot mixtures that include the genetic information for the designed proteins, the template DNA, RNA polymerase, ribosomes, and cofactors for transcription and translation.

Designing genetically encoded scaffold structures for the controlled spatial arrangement of multiple functionalities on the 10– to 100-nm scale is an important unmet challenge for synthetic biology and bio-nanotechnology. Such objects would be of interest for fundamental research, medicine, and industrial biotechnology. For example, controlling the spatial arrangement of molecules in cells could help researchers to explore the connection between spatial order and function on multiple scales, in particular in the context of gene regulation (1) and in signaling (2, 3). Therapeutically active assemblies could be envisioned whose function is logically connected to cellular processes. Scaffolding of multiple enzymes into a common three-dimensional (3D) framework could also facilitate the production of desired compounds more efficiently in engineered cells or cell extracts (4).

Natural and designed proteins provide access to a vast library of functional modules for molecular recognition and enzymatic catalysis (5–8), and proteins can also serve as hubs for forming higher-order assemblies (9, 10). DNA nanotechnology (11), on the other hand, offers the possibility to construct arbitrary shapes (12–17), periodic arrays (18–22), and mechanisms (22–24) with overall or repeat unit dimensions ranging from tens to hundreds of nanometers. Integrating protein-based functionalities into designed DNA-based structural scaffolds could thus provide an attractive approach toward creating multifunctional assemblies with user-defined shapes (25–28). However, although proteins can easily be genetically encoded and produced in vivo, the encoding and assembly of complex DNA-based structures inside cells face major hurdles. For example, DNA nanostructures typically cannot fold without prior annealing at elevated temperatures (29), and such objects assemble from DNA single strands, which are not easily available inside living cells. Designed DNA-based nanostructures could potentially be assembled inside cells through cotranscriptional folding (30); however, it is not yet clear how additional functionalities could be integrated into such RNA scaffolds.

Design of staple proteins

We reimagined the scaffolded DNA origami design principle (12) from DNA nanotechnology. In contrast to DNA origami, which uses DNA single strands as building blocks, our approach uses double-helical DNA and proteins. In our method, a continuous double-helical DNA template with user-defined length assumes the role of a scaffold that is folded into a user-defined 3D arrangement by a set of designed staple proteins (Fig. 1A; see also fig. S1) (31). Each staple protein is tailored to specifically recognize two different double-helical DNA sequences with user-defined separation on the double-helical DNA template. The staple proteins can carry additional functionalities. Inspired by the antiparallel structures of the N-terminal domain and the choice of 21-bp folding frame for translation, we linked the C terminus can be genetically encoded in a single open reading frame for translation, we linked the C terminus with the N terminus of another TAL (Fig. 1B, top inset). Given the absence of sufficiently detailed structural information about the N-terminal TAL domain and the properties of the C-terminal exit point, we had to design and experimentally test several different candidate peptide linker sequences (fig. S4A). On the basis of expression yield in Escherichia coli, solubility, and the ability to link two independent DNA double strands (fig. S4B), we selected one linker sequence out of 11 candidate sequences and used this linker in all staple proteins in this work. We also fused a GFP domain to the N terminus of the double-TAL staple protein. The fluorescence of GFP served to facilitate purification and analysis, and also acted as a placeholder for another protein domain with custom functionality.

To illustrate the shape of a double-TAL staple protein thus constructed and bound to two template double-helical DNA domains, we created a structural model consisting of two antiparallel DNA-bound TAL molecules (34) and calculated electron density projections from various directions (Fig. 1, C and D). The M-shaped projections agree well with the experimental realization of the antiparallel staple-protein connection in the context of larger DNA-protein hybrid structures as observed in electron microscopy experiments (Fig. 1, C and D).

Junction spacing rules

Having established the elementary double-TAL staple protein, we proceeded to identify a set of

*Corresponding author. Email: dietz@tum.de
design rules for creating more complex structural frameworks based on multiple double-TAL staple proteins and a long double-helical DNA template. A necessary step toward creating such more complex objects is the ability to connect two parallel helices by multiple staple proteins. Because of the helicity of B-form DNA, shifting one binding site relative to another on the same helix causes a rotation of one staple protein junction relative to the other. To obtain planarity, it is thus important to find suitable relative positions of the two binding sites on each DNA helix, which will depend on the length of the double-helical spacer between the binding sites (Fig. 1B, bottom right inset). Because the binding sequence for a TAL protein can be contained in either of the two strands of a DNA double-helical domain, the TAL proteins can bind in different orientations. The relative orientation of the two adjacent staple proteins will affect the choice of suitable spacing between adjacent binding sites. In total, there are six unique relative orientations of two adjacent staple proteins (Fig. 2A to D). In DNA nanotechnology, junction spacing rules are derived from the geometry of B-form DNA (12, 14, 32). Here, the detailed geometrical properties of the staple protein connection were not available to us. For the configuration in which the adjacent TAL staple proteins have the same orientation (Fig. 2A), planarity is achieved by simply spacing the binding sites in integer multiples of a single B-form DNA helical turn. For the other five possible configurations, we had to determine empirically the junction spacing rules. On the basis of TEM imaging of a set of test structures (see fig. S5), we arrived at the spacing rules shown in Fig. 2.
The double-crossover motif in Fig. 2A features two staple proteins in the same orientation. The image obtained from the resulting structure resembles two copies of the antiparallel crossover motif from Fig. 1D, showing an MM-shaped signature on the top DNA double strand. By contrast, the motif in Fig. 2B features two staple proteins in opposite orientations. The change in orientation is reflected in the resulting image, which now shows an MW-shaped signature on the top DNA double strand. Based on the spacer lengths, we again constructed structural models and calculated electron density projections (Fig. 2, A and B). The projections agree well with the experimental image data and reproduce the subtle differences in the data from the different orientations.

The junction spacing rules shown in Fig. 2, A to D, enable construction of larger DNA-protein hybrid nanostructures, which may be achieved by combining multiple double crossover motifs such that staple proteins with the same orientations are spaced with distances corresponding to integer multiples of 21 bp. For example, the configuration in Fig. 2A may be simply repeated along the helical direction to form a filamentous two-helix bundle. Alternatively, the configuration in Fig. 2B may be used for the same purpose by alternating the spacer lengths (Fig. 2E).

To see whether some configurations may be preferable over others for creating larger hybrid objects, we constructed four distinct two-helix bundles using a set of 12 different staple proteins. For each two-helix bundle variant, we designed a custom DNA template sequence containing 24 unique recognition sites. We found that two-helix bundles built using repetitions of the double-crossover motifs in Fig. 2, A and D, tended to form multistranded higher-order structures, whereas the bundles based on the motifs in Fig. 2, B and C, remained mostly monomeric (fig. S6). We speculate that the highly repetitive patterns of TAL protein surface features can promote avidity effects in protein-protein interactions. This effect is strongly enhanced when all binding sites on one DNA double helix point into the same direction, as is the case for the double-crossover motifs in Fig. 2, A and D. Hence, to minimize the undesired repetitive interaction motifs, for further designs we used the alternating staple protein orientations from Fig. 2, B and C. To illustrate the site selectivity of our staple proteins, we incubated the template DNA used in the two-helix bundle shown in Fig. 2E with distinct subsets of the 12 staple proteins. The TEM data obtained from the resulting objects revealed binding of the staple proteins in the expected patterns (Fig. 2, F and G).

**Structural motifs**

Having established the fundamental design rules, we explored the possibility of creating structural modules such as custom curvatures, corners, and vertices. We designed a series of template DNA sequences containing 18 unique recognition sites. We found that two-helix bundles built using repetitions of the double-crossover motifs in Fig. 2A may be simply repeated along the helical direction to form a filamentous two-helix bundle. Alternatively, the configuration in Fig. 2B may be used for the same purpose by alternating the spacer lengths (Fig. 2E).

To see whether some configurations may be preferable over others for creating larger hybrid objects, we constructed four distinct two-helix bundles using a set of 12 different staple proteins. For each two-helix bundle variant, we designed a custom DNA template sequence containing 24 unique recognition sites. We found that two-helix bundles built using repetitions of the double-crossover motifs in Fig. 2A and D, tended to form multistranded higher-order structures, whereas the bundles based on the motifs in Fig. 2, B and C, remained mostly monomeric (fig. S6). We speculate that the highly repetitive patterns of TAL protein surface features can promote avidity effects in protein-protein interactions. This effect is strongly enhanced when all binding sites on one DNA double helix point into the same direction, as is the case for the double-crossover motifs in Fig. 2A and D. Hence, to minimize the undesired repetitive interaction motifs, for further designs we used the alternating staple protein orientations from Fig. 2, B and C. To illustrate the site selectivity of our staple proteins, we incubated the template DNA used in the two-helix bundle shown in Fig. 2E with distinct subsets of the 12 staple proteins. The TEM data obtained from the resulting objects revealed binding of the staple proteins in the expected patterns (Fig. 2, F and G).
staple proteins could be recognized. For the square object (Fig. 3F), one corner could be resolved in greater detail, revealing aspects of the staple proteins as well as of the two DNA template strands, which agreed with our design. The branched objects (Fig. 3, D, E, and G) were generally more flexible, but the topology reflected in the single-particle TEM images agreed with the designs. The heterogeneous appearance of the four-armed cross particles (Fig. 3G) may be attributed to collapse of nonplanar tetrahedral conformations onto the 2D TEM imaging support.

Together, the objects in Fig. 3 constitute a set of design motifs that could be combined to create larger, more complex DNA-protein hybrid nanostructures. We also studied the objects in agarose gel-electrophoretic mobility assays (Fig. 3C, bottom, and fig. S6E). For example, the closed ring migrated in reasonably well-defined bands, and we could also purify the rings from such gels by physical extraction (Fig. 3C, bottom). The leading band corresponded to ring monomers; a slower second band corresponded to a dimeric species in which two template DNA molecules were present.
appeared to be connected by the staple proteins, as seen by TEM imaging.

**Multilayer structures**

The objects with topological closure constraint appeared already relatively stiff. Alternatively, enhanced structural rigidity of target objects can also be achieved by connecting multiple layers of parallel helices. To test for the possibility of creating staple protein connections that are compatible with 2D helix packing, we designed a DNA template sequence for a four-helix bundle with a 2 × 2 helix cross section (Fig. 4A). To create the necessary staple protein connections compatible with linkages pointing to multiple neighboring helices, we modified the spacer design configuration from Fig. 2B. To prevent steric conflicts from packing, we increased the overall distance between staple proteins. Additionally, we also shifted every second binding site to induce a relative orientation of adjacent crossovers of ~90° based on B-form DNA geometry (Fig. 4A).

The four-helix bundle thus designed successfully self-assembled at room temperature from a mixture containing the designed template DNA and the 12 unique staple proteins introduced above. TEM imaging revealed rod-like particles (Fig. 4C) with the expected pattern of six segments produced by the transmission projection of the double layer of TAL staple proteins, indicating that the structures folded as designed. As expected, the four-helix DNA-protein rods exhibited greater shape homogeneity relative to the two-helix objects lacking a closure constraint, which we attribute to the desired greater bending stiffness caused by the multilayer design.

Classifying and averaging individual particle micrographs of the four-helix bundle revealed a number of dominant transmission views with distinct thickness of the rod particles in the image plane (Fig. 4C). To understand the appearance of the particles, we constructed a structural model for a segment of our four-helix bundle based on the rhombohedral crystal packing of TAL effectors, as seen in x-ray diffraction experiments (34) and computed electron density projections (Fig. 4D). The computed projections reproduced the thickness variation that we observed in our TEM data. We expect that the flexible linkers in the staple proteins allow the structure to relax into a conformation in which protein-protein interactions also are optimized. Given the correspondence of our TEM data with the computed projections (Fig. 4C versus Fig. 4D), we consider it likely that the four-helix bundle adopts a configuration that resembles the crystal packing of TAL effectors, leading to a diamond-shaped cross section in the helical direction of the four-helix bundle.

Considering also the pronounced tendency to form higher-order structures when using repetitive TAL protein orientations that we observed in several single-layer structures (see fig. S6), we thus conclude that the TAL staple proteins provide additional interaction surfaces that further stabilize the DNA-protein hybrid nanostructures internally, beyond the rotationally unconstrained peptide linker in each staple protein. However, although the packing of the helices appears to be governed by optimizing protein-protein interactions, the global shape of the four-helix bundle is controlled by the designed sequence of the double-helical DNA template. To demonstrate this fact, we prepared a curved variant of the four-helix bundle (Fig. 4E). The curvature was achieved by inserting additional base pairs between adjacent staple protein binding sites in two out of the four helices. TEM imaging confirmed that the curved four-helix bundle self-assembled successfully from the template DNA thus modified and the same set of 12 staple proteins that was used for the straight variant.

**Genetically encoded self-assembly**

Finally, we tested the possibility for self-assembly at physiological conditions of our DNA-protein hybrid nanostructures starting from genetic expression (Fig. 5). To this end, we used an in vitro expression one-pot mixture comprising RNA polymerase, ribosomes, and the various auxiliary cofactors that are necessary for transcription and translation, such as nucleotides, amino acids, transfer RNAs (tRNAs), tRNA transferases, etc. To this mixture, we added a template strand for a hybrid nanostructure and 12 different plasmids featuring the genes for 12 different double-TAL staple proteins. After 6 hours of incubation at 25°C,
intact self-assembled DNA-protein hybrid nanostructures emerged and could be discerned side-by-side with ribosomes and other components in TEM images acquired from the self-assembly reaction mixtures. To verify (as in Fig. 2, F and G) the site specificity of staple protein binding, we used only 10 of the 12 double-TAL staple genes in combination with the template for the hybrid ring. TEM imaging revealed that the structures formed in this reaction mixture, as expected, contained gaps at the binding sites for the two staple proteins that were excluded (Fig. 5D, red arrows).

Discussion

In this study, we developed design rules for creating custom single-layer and multilayer DNA-protein hybrid nanostructures; using these rules, we successfully self-assembled 18 distinct objects. We have observed a tendency for aggregation that we attribute to protein-protein interactions. In the future, the rational reengineering of the TAL surfaces using computational protein design (39) may enable suppression of unwanted interactions in TAL repeats at the outside of assembled objects while preserving those interactions that stabilize the objects internally. We have used linear polymerase chain reaction (PCR) products as template DNA to produce the majority of the objects shown in Figs. 1 to 5, but circular plasmids may be used just as well (Fig. 5C, fig. S6, E and F, and fig. S7). Thus, all of the components needed for the assembly of our hybrid nanostructures can be encoded genetically. Moreover, the assembly of our structures proceeds upon simple mixing at room temperature at near-physiological buffer conditions, which allows for self-assembly after transcription of the genes, as we demonstrated in Fig. 5. In vivo assembly of our DNA-protein hybrid nanostructures appears thus a realistic possibility.

Regarding the integration of multiple functionalities, additional protein domains may be readily fused to the staple proteins by conventional protein engineering methods to arrange these groups in a user-defined 3D spatial arrangement. In our experiments, each staple protein already carried a placeholder for such a protein domain with user-defined functionality, as represented by GFP domains fused to the N termini of the double-TAL staple proteins.

In our experiments, we created a diversity of structures by designing custom DNA template sequences while reusing a set of 12 unique staple proteins. Redesigning the template DNA sequences while reusing, for example, our set of 12 staple proteins is convenient in that it only...
requires gene synthesis. However, given the programmability of TAL proteins, a virtually unlimited number of staple proteins could be created. Designing additional double-TAL staple proteins may be implemented with efficient TAL cloning methods (40, 41), which is also the approach that we adopted initially to create the 12 unique double-TAL staple proteins. Designing a set of custom staple proteins in order to fold a given target DNA double strand into a desired shape could be of interest, for example, as a means of rationally creating custom structures and loops in cellular genomes to dissect the influence of genome looping on gene regulation (7).

Many current applications of bottom-up nanotechnology, and DNA nanotechnology in particular, rely on positioning or displaying of a few target features, often proteins, with multiple functional proteins by terminal fusions to the staple proteins in order to fold a given target DNA double strand into a desired shape could be of interest, for example, as a means of rationally creating custom structures and loops in cellular genomes to dissect the influence of genome looping on gene regulation (7).

**Assembly of DNA-protein hybrid nanostructures**

All template sequences were synthesized by Eurofins Genomics (Ebersberg, Germany) and delivered in a standard plasmid backbone (pEX-A2 or pEX-K4). Linear dsDNA templates were produced by PCR using primers that bind in the plasmid backbone yielding template strands with approximately 100- to 130-bp overhangs before the first and after the last binding sequence, respectively. Assembly of hybrid nanostructures was performed by mixing the template DNA at concentrations between 5 and 20 μM with a 4-fold excess per staple protein over template DNA using a solution of pre-mixed staple proteins, and incubating for one to four hours at room temperature.

**TEM imaging and image processing**

Samples were diluted to final concentrations between 0.5 and 2 μM and adsorbed on glow-discharged Formvar-supported carbon-coated Cu400 TEM grids (Science Services, Munich) and stained using a 2% aqueous uranyl formate solution containing 25 mM NaOH. Imaging was performed using either a Philips CM100 EM operated at 100 kV or a FEI Tecnai Spirit operated at 120 kV. Reference-free class averaging was performed using Xmipp 3.0 (49). Models based on the crystal structure were constructed using UCSF Chimera (50). TEM micrographs were subjected to bilinear downsampling (ImageJ), high-pass filtering to remove long-range staining gradients, sharpening to enhance particle edges, and contrast auto-leveling (Adobe Photoshop CS5).

**Gel electrophoresis of DNA-protein hybrid nanostructures**

Hybrid nanostructures were assembled as described above and electrophoresed in a 1.5% agarose gel, in an ice bath. After scanning the gel in a Typhoon FLA 9500 laser scanner (GE), the desired bands were cut out and applied to a 0.45 μm spin filter. After centrifugation for 5 min at 5000 rcf (4°C) the filtrate was immediately applied to Formvar-supported carbon-coated Cu400 TEM grids and stained as described above.

**One pot in vitro expression and self-assembly of DNA-protein hybrid nanostructures**

For in vitro expression, the PURExpress system (New England Biolabs) was used. Reactions with a final volume of 25 μl contained 10 μl solution A, 7.5 μl solution B, 1 μl murine RNase inhibitor (New England Biolabs), 1080 ng plasmids (12 plasmids @ 90 ng per plasmid), and 5 μM template DNA. Reaction mixtures were incubated at 25°C for 6-8 hours, diluted 1,000-10,000 fold and immediately applied to TEM grids and stained as described above.

**REFERENCES AND NOTES**


Praetorius et al., Science 355, eaam5488 (2017) 24 March 2017

7 of 8

Downloaded from http://science.sciencemag.org on March 23, 2017


31. See supplementary materials.


ACKNOWLEDGMENTS

We thank S. Ständer, M. Homemann, and J. Valtin for technical assistance. K. Wagenbauer for help with TEM imaging, and T. Gerling and E. Stahl for discussions. Supported by European Research Council starting grant 295270 (H.D.), Bundesministerium für Bildung und Forschung grant 031 A 458, and the Deutsche Forschungsgemeinschaft through grants provided within the Gottfried-Wilhelm-Leibniz Program and the Excellence Cluster CIPSM (Center for Integrated Protein Science Munich). All data are reported in the main text and supplement.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/355/6331/eaam5488/suppl/DC1

Materials and Methods
Figs. S1 to S22
 Tables S1 to S15
 hybrid_structure_generator.xlsx file

7 December 2016; accepted 13 February 2017
10.1126/science.aam5488
Editor's Summary

**Protein-folded DNA nanostructures**

A wide variety of DNA nanostructures have been assembled by folding long DNA single strands with short DNA staples. However, such structures typically need annealing at elevated temperatures in order to form. To accommodate the formation of such structures in living cells, Praetorius and Dietz developed an approach in which custom protein staples based on transcription activator–like effector proteins fold double-stranded DNA templates (see the Perspective by Douglas). The structures folded into user-defined geometric shapes on the scale of tens to hundreds of nanometers. These nanostructures could self-assemble at room temperature in physiological buffers.

*Science, this issue* p. eaam5488; see also p. 1261