This is a repository copy of *Directed Assembly of 3-nm-long RecA Nucleoprotein Filaments on Double-Stranded DNA with Nanometer Resolution*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/78726/

ARTICLE

Directed Assembly of 3-nm-long RecA Nucleoprotein Filaments on Double-Stranded DNA with Nanometer Resolution

Rajan Sharma, A. Giles Davies, and Christoph Wälti*

School of Electronic and Electrical Engineering, University of Leeds, Leeds LS2 9JT, U.K.

ABSTRACT Protein-mediated self-assembly is arguably one of the most promising routes for building complex molecular nanostructures. Here, we report a molecular self-assembly technique that allows programmable, site-specific patterning of double-stranded DNA scaffolds, at a single-base resolution, by 3-nm-long RecA-based nucleoprotein filaments. RecA proteins bind to single-stranded DNA to form nucleoprotein filaments. These can self-assemble onto a double-stranded DNA scaffold at a region homologous to the nucleoprotein's single-stranded DNA sequence. We demonstrate that nucleoprotein filaments can be formed from single-stranded DNA molecules ranging in length from 60 nucleotides down to just 6 nucleotides, and these can be assembled site-specifically onto a model DNA scaffold both at the end of the scaffold and away from the end. In both cases, successful site-specific self-assembly is demonstrated even for the smallest nucleoprotein filaments, which are just 3 nm long, comprise only two monomers of RecA, and cover less than one helical turn of the double-stranded DNA scaffold. Finally, we demonstrate that the RecA-mediated assembly process is highly site-specific and that the filaments indeed bind only to the homologous region of the DNA scaffold, leaving the neighboring scaffold exposed.
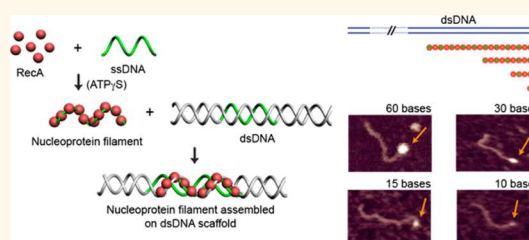
KEYWORDS: self-assembly · nanomaterials · RecA protein · triple-strand assembly · homologous recombination

The synthesis of functional nanostructured materials, and in particular materials that are heterogeneously structured on the nanoscale, is one of the greatest challenges of nanosciences and technologies. Owing to the significant advancements of the field in recent years, a number of novel, high-performance materials for a vast range of applications have emerged. However, these materials are generally homogeneous at the nanoscale and hence do not offer functional or structural variety at these length scales. The fabrication of materials that require highly heterogeneous nanoscale structures, for example, where the material requires distinct functionality at different locations, is still a significant challenge. The synthesis of such materials will almost certainly be limited to appropriate bottom-up approaches, where the material is self-assembled from a number of building blocks. A very promising approach for the directed assembly of nanoscale components is the exploitation of molecular self-assembly, taking advantage of the specific lock-and-key recognition properties characteristic of many biological systems.

Molecular self-assembly is inherent to a number of biological molecules and has been exploited widely over the past few years. Molecules such as DNA have been explored for the fabrication of two- and three-dimensional self-assembled DNA nanostructures.[1−4] Such structures have been used as scaffolds to template and assemble complex molecular-scale objects, for example, to construct arrays of metallic nanoparticles to form molecular-scale nanowires.[5,6]

However, these techniques generally do not offer flexible site-specific assembly and do not allow the scaffold to be patterned programmably with high spatial accuracy. To address this, alternative patterning mechanisms using recombinase A (RecA)-based

ARTICLE

DNA nucleoprotein filaments have been explored.[7−10] RecA is the major protein responsible for the homologous recombination in *Escherichia coli*,[11] a protein-mediated process by which a single-stranded (ss)DNA oligonucleotide is paired through self-assembly[12] and eventually exchanged with the homologous region of one of the strands of a double-stranded (ds)DNA molecule, *i.e.*, the region with the same or a very similar base sequence on the dsDNA.[13,14] RecA-based DNA nucleoprotein filaments designed to self-assemble at sequence-specific positions on dsDNA scaffold molecules are created by polymerizing RecA monomers around the appropriate homologous ssDNA, with one RecA monomer per three nucleotides.[15]

Examples of this approach include the use of nucleoprotein filaments to mask a dsDNA scaffold locally from a subsequent metalization process[10] and to create both a metal binding site and a region unsusceptible to metal binding on the same DNA molecule.[9] In these experiments, RecA nucleoprotein filaments with lengths well in excess of 100 nm were employed. We have previously shown that programmable and concurrent patterning with RecA-based nucleoprotein filaments at different locations on the same DNA scaffold is possible.[7] However, for directed and programmable assembly at the ultimate molecular scale, mechanisms offering patterning with feature sizes and resolutions at a scale significantly smaller than 10 nm are needed, hence requiring RecA nucleoprotein filaments created from constituent ssDNA molecules significantly shorter than 20 bases.

Previous studies have shown complete RecA-based nucleoprotein filament formation only for constituent ssDNA longer than 15 bases[16−18] and that such nucleoprotein filaments can bind to circular but not linear dsDNA molecules with a homologous region of at least 8 nucleotide bases.[19,20] The smallest nucleoprotein filament successfully assembled onto a linear dsDNA molecule was created using a 18-nucleotide-long ssDNA molecule,[21] and it was estimated using FRET that a filament larger than 14 bases is required to search for the region of homology on the dsDNA scaffold.[22] In contrast, another study reported that nucleoprotein filaments formed from 30- to 50-base-long ssDNA were required to enable self-assembly onto circular dsDNA molecules, with the assembly being less efficient when linear dsDNA molecules were used as the scaffold.[23] These findings were largely confirmed by a number of other studies that primarily employed either circular ssDNA or circular dsDNA scaffolds.[24,25] However, the formation of nucleoprotein filaments from ssDNA only a few nucleotides long and their subsequent self-assembly onto DNA scaffolds with molecular precision, an important prerequisite for the fabrication of advanced functional nanoscale materials, has not been demonstrated to date.

Here we demonstrate that the size of the nucleoprotein filaments can be reduced considerably while at the same time maintaining their highly specific and programmable assembly properties. We show that nucleoprotein filaments formed from only 6-nucleotide-long ssDNA molecules (*i.e.*, 3 nm in length) can assemble onto homologous regions on dsDNA scaffolds, covering significantly less than one helical turn of the DNA.

RESULTS AND DISCUSSION

In the presence of a nucleotide cofactor (ATP or nonhydrolyzable ATPγS (adenosine 5′-*O*-(3-thiotriphosphate))), RecA proteins adopt a state with strong affinity for both ssDNA and dsDNA[26] and at low magnesium concentrations (1−2 mM) undergo a slow nucleation process onto the DNA. A helical nucleoprotein filament is then formed in a rapid and highly cooperative extension process with additional RecA monomers adding to the ends of the previously nucleated monomers.[16,27,28] One RecA monomer binds to three nucleotide bases or base-pairs (bp) on the DNA, and the resulting nucleoprotein filaments are about 50% longer than the equivalent B-form DNA, with a pitch of 6.2 RecA monomers.[29]

Kinetic studies of RecA nucleoprotein formation when using ATP as the cofactor have shown the on- and off-rates of the RecA monomer assembly on the 5′ end of the nucleoprotein filament to be $k_{on} = 0.11$ s^{-1} and $k_{off} = 0.12$ s^{-1}, respectively, with the dissociation constant estimated to be on the order of 100 nM.[16] In contrast, for assembly on the 3′ end of the nucleoprotein filament, the dissociation constant was found to be 8 nM, with $k_{on} = 0.18$ s^{-1} and $k_{off} = 0.16$ s^{-1}.[16] The 10-fold difference in the binding rates between the two ends of the nucleoprotein filament dictates the direction of filament growth. The main driver initiating the RecA monomer dissociation is hydrolysis of the ATP, and when ATPγS is used as the cofactor, the filaments are essentially stable.[27]

If nucleoprotein filaments are exposed to dsDNA molecules that share a region of homology, a triple-stranded DNA structure encapsulated in RecA proteins is formed (see Figure 1) at Mg^{2+} concentrations of around 10 mM.[30] This facilitates strand exchange during the homologous DNA recombination process.[31]

The RecA-mediated DNA triple-stand formation is a two-step process. In the first step the nucleoprotein filaments bind nonspecifically to nonhomologous regions on the dsDNA through local destabilization of the dsDNA scaffold to form a synaptic heterologous nucleoprotein filament−dsDNA complex.[22,32] The apparent dissociation constant of this reaction has been estimated to be <10 pM,[32] and the lifetime of the heterologous nucleoprotein filament−dsDNA complex was found to be between 0.5 and 10 s.[32,33] In the second step, the bound nucleoprotein filament then searches for the homologous assembly site on the dsDNA
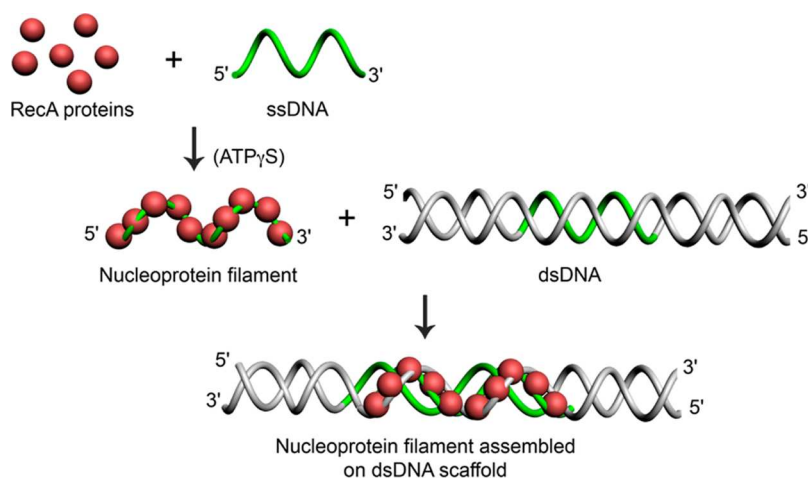
Figure 1. Schematic diagram illustrating the RecA protein-mediated assembly of a nucleoprotein filament onto the region of a dsDNA scaffold with base sequence homologous to the ssDNA from which the nucleoprotein filament is formed.

scaffold. This is significantly slower than the first step and is the rate-limiting rate constant for RecA-mediated directed assembly.[32] The rate of search for homology depends on the length of the constituent ssDNA.[32] For filaments formed from 24-nucleotide-long ssDNA, the rate constant is estimated, using restriction assays, to be on the order of $(2-4) \times 10^{-4}$ s^{-1}.[34] The corresponding diffusion coefficient of the nucleoprotein filament in the search for homology has been found to be approximately 8000 bp^2/s.[33]

Having found the homologous binding site, the dissociation of the nucleoprotein filament from the dsDNA occurs *via* disassembly of the RecA monomers from the filament itself, driven by ATP hydrolysis.[11,25] However, if ATPγS is used as the cofactor, the nucleoprotein filaments are found not to dissociate even after 2 h,[25] providing sufficiently stable nucleoprotein filament–dsDNA complexes for nanoscale patterning applications.

Nucleoprotein filament assembly onto dsDNA scaffolds has been previously demonstrated for filaments made from 60-base-long (approximately 30 nm) homologous ssDNA molecules.[7] To investigate the potential reduction in minimum feature size of the nucleoprotein filaments for programmable nanoscale assembly, a 138 bp dsDNA molecule, approximately 47 nm in length, was designed to serve as a model scaffold. An Apol restriction site (GAATTC) was included in the scaffold, centered 7 bases away from the 3′ end, to enable the investigation of the efficiency of the site-specific assembly of the nucleoprotein filaments onto the scaffold through a subsequent restriction enzyme digestion assay. The restriction assay is illustrated in Figure 2, and the schematic of the design of the 138 bp scaffold is given in Figure 3a. A series of nucleoprotein filaments was created from ssDNA of lengths between 60 and 6 nucleotides, all homologous to the 3′ end of the 138 bp dsDNA scaffold (see Figure 3a), and the ability of these filaments to
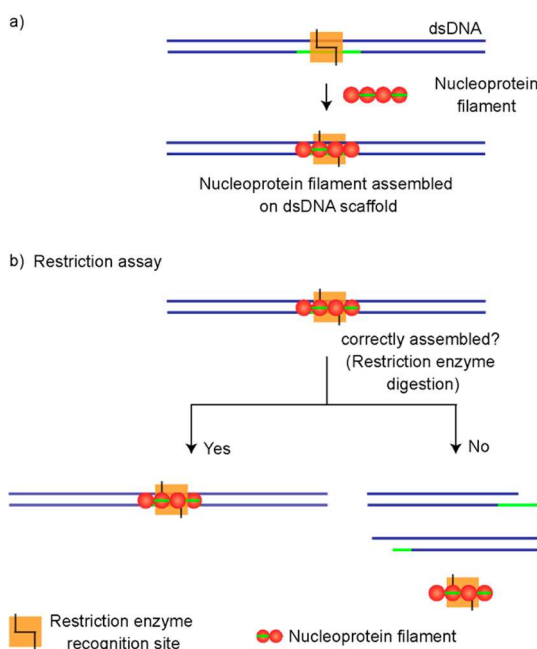

Figure 2. (a) Schematic illustration of RecA protein-mediated self-assembly of a nucleoprotein filament with constituent ssDNA homologous to a region on the dsDNA comprising a restriction enzyme recognition site. (b) A nucleoprotein filament self-assembled onto the correct location on the DNA scaffold masks the restriction enzyme recognition site and thus inhibits enzymatic cleavage of the dsDNA by the enzyme. However, the dsDNA scaffold is digested into shorter fragments if the nucleoprotein filament fails to self-assemble onto the correct location.

assemble onto the dsDNA scaffold assessed. The yield of the assembly process was determined by attempting to digest the dsDNA scaffold with the restriction enzyme Apol following the assembly process; correct nucleoprotein filament assembly protects the scaffold from digestion at the restriction site (see Figure 2).

The filaments were then assembled onto the scaffold in solution by keeping the ratio of nucleoprotein filaments to dsDNA scaffold molecules constant at 40:1 to ensure a high patterning yield is obtained, at least
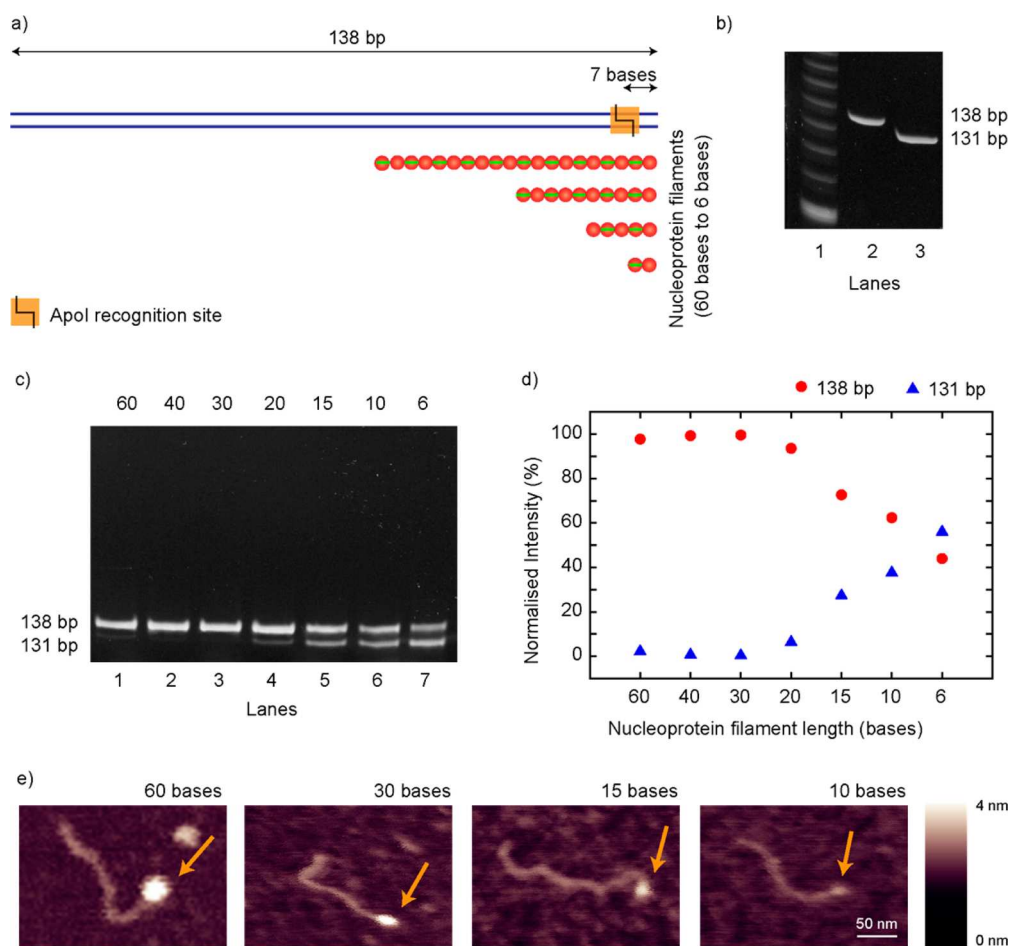
Figure 3. (a) Schematic diagram of a 138 bp DNA scaffold showing the location of the ApoI recognition site (centered 7 bases away from the end). The regions where the different nucleoprotein filaments made from ssDNA of lengths between 60 and 6 bases bind onto the dsDNA scaffold are also indicated. Upon assembly, the nucleoprotein filaments protect the ApoI recognition site of the dsDNA from subsequent enzymatic digestion. (b) 13% PAGE gel (19:1, 120 V, 5 h) showing the 138 bp DNA scaffold (lane 2), the 138 bp DNA scaffold digested with ApoI in the absence of any RecA protein resulting in 131 bp DNA (lane 3), and a molecular weight marker (lane 1). (c) 13% PAGE gel (19:1, 120 V, 5 h) showing the effect of the length of the ssDNA molecule used to create the nucleoprotein filament on the efficiency of assembly. Lanes 1–7 show the 138 bp DNA scaffolds digested with ApoI while protecting the ApoI site with nucleoprotein filaments of different length. The length of the ssDNA used to create the nucleoprotein filament is indicated above the lanes and ranges from 60 to 6 nucleotides. (d) Normalized intensities of the 138 bp and 131 bp bands in lanes 1–7 of panel (c) versus nucleoprotein filament length (60 to 6 bases). (e) AFM images of nucleoprotein filaments formed from ssDNA homologous to one end of the scaffold of lengths 60, 30, 15, and 10 nucleotides, respectively, self-assembled onto a 711 bp dsDNA scaffold. The nucleoprotein filaments are indicated by arrows.

for the longest filaments, which have already been demonstrated to assemble at high yields at this ratio.[7] The yield of assembly was then determined indirectly using the restriction digestion assay; that is, the assembled triple-stranded complexes were digested with ApoI, treated with proteinase-K to remove any proteins still bound to the DNA scaffold, and then run on a polyacrylamide gel for band separation and inspection. We note that in the absence of any nucleoprotein filaments the ApoI restriction enzyme digests the 138 bp DNA at the ApoI site, resulting in two smaller fragments: 131 bp and 7 bp (lane 3, Figure 3b). To ensure sufficient separation between the undigested DNA (138 bp) and the longer one of the two fragments resulting from the digestion (131 bp), the PAGE gel was run for 5 h, which resulted in the short

fragment (7 bp) running off the gel. The results of the restriction assays for all filaments are shown in Figure 3c.

For the DNA scaffold that was protected through the site-specific assembly of the nucleoprotein filament formed from a 60-nucleotide-long ssDNA molecule, only one band, corresponding to the 138 bp scaffold, can be seen on the gel (Figure 3c), indicating that the DNA scaffold was protected with high yield and consequently that the yield of the assembly process was high. The same findings were obtained for nucleoprotein filaments formed from 40- and 30-nucleotide-long ssDNA. When reducing the length of the ssDNA encapsulated in the nucleoprotein filament and thus the length of the nucleoprotein filaments further, successful site-specific self-assembly can still be observed,

albeit with reduced yield. This can be seen from the appearance of a second band, corresponding to 131 bp dsDNA, which indicates the presence of digested scaffold molecules owing to the reduced yield of protection from digestion. Importantly, even for the shortest nucleoprotein filament, which is formed from a ssDNA molecule of just 6 nucleotides long and hence 3 nm in length, successful site-specific assembly was observed (lane 7, Figure 3c).

The yield of the protection and hence the yield of site-specific assembly can be estimated from the normalized intensity of the 138 bp band, *i.e.*, the band that corresponds to the undigested protected fraction of DNA scaffolds. From the intensity plot (Figure 3d) it can be seen that for long nucleoprotein filaments yields well in excess of 95% can be achieved and that even for nucleoprotein filaments as short as 5 and 3 nm yields in excess of 60% and 40% are obtained, respectively. We note that the 3-nm-long nucleoprotein filaments are assembled from only two RecA monomers and that the nucleoprotein filament−dsDNA complexes formed from even the shortest filaments in this study were stable for at least 30 min, in agreement with previous findings.[34]

Figure 3e shows AFM images of 711 bp dsDNA scaffolds onto which nucleoprotein filaments of various lengths have been assembled, ranging from nucleoprotein filaments created using 60- down to 10-nucleotide-long ssDNA. The 711 bp scaffold is similar to the 138 bp scaffold used for the restriction assay with the 3′ end identical to that of the 138 bp scaffold, but the 5′ end extended significantly for the AFM study. A clear difference in size of the nucleoprotein filaments (indicated by arrows) assembled onto the DNA scaffold can be seen for nucleoprotein filaments decreasing in length. The length of the ssDNA used to create the nucleoprotein filament is indicated above each image. Again, successful assembly can be seen even for nanoscale nucleoprotein filaments.

These results demonstrate that even 3-nm-long nucleoprotein filaments can be self-assembled successfully onto DNA scaffolds, which is a significant improvement over previously established limits, where successful assembly of RecA nucleoprotein filaments was observed on circular dsDNA only with the filaments longer than 15 bases and on linear dsDNA scaffolds with at least 18-base-long nucleoprotein filaments. This brings this technique well into the molecular and nanoscale regime, opening up a range of previously inaccessible avenues for programmable site-specific self-assembly.

The RecA-facilitated self-assembly of nucleoprotein filaments onto dsDNA scaffolds has previously been shown, using nucleoprotein filaments of 30 bases or longer, to be dependent on the distance of the assembly location from the end of the scaffold.[35] To investigate the effect of the position of the homologous

region on the dsDNA scaffold on the assembly of nucleoprotein filament−dsDNA complexes, a new 189 bp dsDNA model scaffold was designed as shown in Figure 4a. The new scaffold is similar to the 138 bp scaffold, but extended on the 3′ end by 51 bp such that the assembly region is now located 58 bases away from the 3′ end. Identical nucleoprotein filaments to the ones used previously (Figure 3) were then self-assembled onto the new scaffold in a similar way to that described above. A restriction assay was again used to estimate the assembly yields, and the results are shown in Figure 4c,d.

For nucleoprotein filaments formed from a 60-nucleotide-long ssDNA molecule, a bright prominent band corresponding to the 189 bp undigested scaffold is observed, with only a faint band corresponding to the digestion product (see Figure 4c, lane 1). This shows that although the assembly site has been moved away from the end of the scaffold, the assembly yield is still in excess of 80%. Upon decreasing the length of the nucleoprotein filament by reducing the size of the constituent ssDNA molecule, a reduction in the 189 bp:131 bp intensity ratio is observed, which indicates a reduction in assembly yield with decreasing filament length. This is similar to the situation for assembly at the end of the scaffold. Figure 4d shows the normalized intensities of the bands, and the normalized intensity of the 189 bp can be used as a means to estimate the assembly yield. However, from this figure it can be seen that even 3-nm-long nucleoprotein filaments, comprising only two RecA monomers, still self-assemble site-specifically, but now at a location away from the end of the dsDNA scaffold. We note that for the same length of constituent homologous ssDNA the assembly of nucleoprotein filaments on the dsDNA scaffold is found to be less efficient when the assembly site was moved away from the end of the scaffold (see Figures 3d and 4d). This is in agreement with our previous findings where, for nucleoprotein filaments formed from 60-base-long ssDNA, an increase of the nucleoprotein filament to dsDNA scaffold ratio was shown to compensate the reduction in assembly efficiency upon moving the assembly site away from the end of the scaffold.[7] Figure 4e shows AFM images of nucleoprotein assembly onto a 926 bp dsDNA scaffold, which is similar to the 189 bp scaffold but extended on both the 5′ and 3′ ends for AFM investigations. Self-assembled complexes of filaments formed from 60-nucleotide-long and 10-nucleotide-long ssDNA onto dsDNA are shown in the upper and lower panel, respectively. These images clearly show the successful site-specific assembly of very short nucleoprotein filaments onto dsDNA scaffolds.

In order to demonstrate the site-specificity and site-selectivity of the RecA-mediated self-assembly process, in particular for the short 3-nm-long nucleoprotein filaments, three model scaffolds, 130 bp-1,
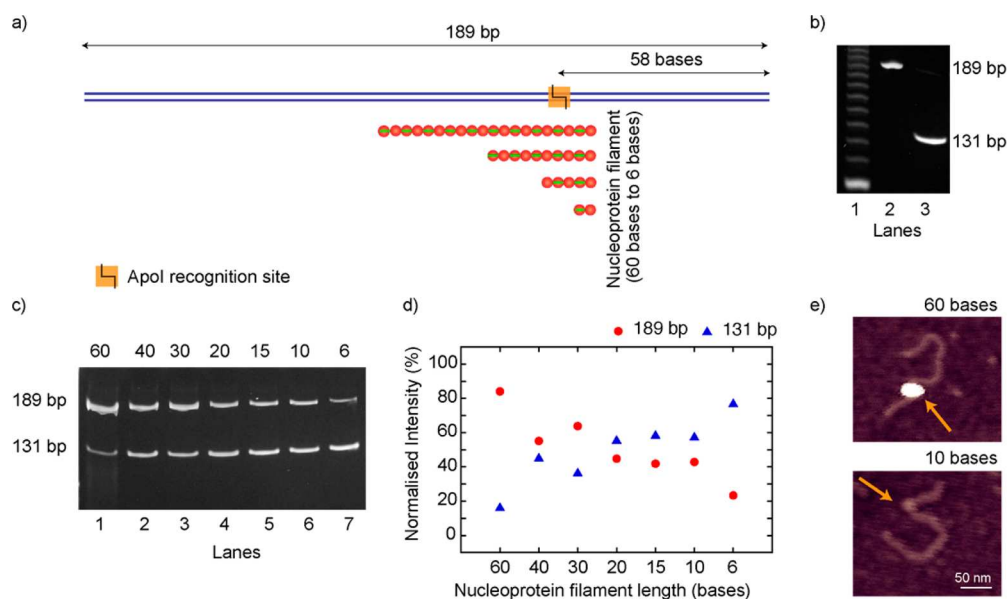
Figure 4. (a) Schematic diagram of a 189 bp DNA scaffold showing the location of the ApoI recognition site (centered 58 bases away from the end) and the regions where the different nucleoprotein filaments with constituent ssDNA of lengths between 60 and 6 bases bind onto the dsDNA scaffold. Upon assembly, the nucleoprotein filaments cover and therefore protect the ApoI recognition site of the dsDNA from subsequent enzymatic digestion. (b) 15% PAGE gel (19:1, 80 V, 9 h). Lane 1: molecular weight marker; lane 2: 189 bp DNA scaffold; lane 3: 189 bp DNA scaffold digested with ApoI in the absence of any RecA-mediated protection resulting in 131 bp DNA. (c) 15% PAGE gel (19:1, 80 V, 9 h). Lanes 1−7: 189 bp DNA scaffolds digested with ApoI while protecting the ApoI site with nucleoprotein filaments of different length. The length of the ssDNA used to create the nucleoprotein filament is indicated above the lanes and ranges from 60 to 6 nucleotides. (d) Normalized intensities of the 189 bp and 131 bp bands in lanes 1−7 of panel (c) versus nucleoprotein filament length. (e) AFM images of 926 bp dsDNA scaffolds patterned with nucleoprotein filaments (indicated by arrows) formed from ssDNA homologous to a region 215 bp away from one end of the scaffold of lengths 60 and 10 nucleotides, respectively.

130 bp-2, and 130 bp-3, were designed (see Figure 5a). The 130 bp-1 scaffold is similar to the original 138 bp scaffold (see Figure 3a) and has an ApoI restriction site (site-1) centered 7 bp away from the 3′ end, but in addition has a second ApoI site (site-2) centered 17 bp away from the 5′ end. Scaffolds 130 bp-2 and 130 bp-3 are variants of 130 bp-1 with different 3′ ends. The three scaffolds were generated by an overlap extension PCR method, to place the ApoI site-1 at a position centered 7 bases, 12 bases, and 24 bases away from the 3′ end in the 130 bp-1, 130 bp-2, and 130 bp-3 scaffold, respectively. The position of the ApoI site-2 remained the same for all three scaffolds.

Nucleoprotein filaments were generated using 6-base-long ssDNA molecules homologous to the 3′ end of the three scaffolds, as indicated in Figure 5a. For the 130 bp-1 scaffold, the assembly location of the nucleoprotein filament partially overlaps the ApoI site-1, but it is 6 bases away from the center of the ApoI site-1 in the 130 bp-2 scaffold and 18 bases away from the center of the ApoI site-1 of the 130 bp-3 scaffold. 3-nm-long nucleoprotein filaments were assembled onto all three scaffolds in a similar way to that discussed above. The yield of the assembly was estimated with a restriction assay, and the results are shown in Figure 5b.

For the 130 bp-1 scaffold, the 3-nm-long nucleoprotein filament is expected to assemble partly across

the ApoI site-1, therefore preventing digestion of the scaffold at the ApoI site-1. In contrast, the filaments are not expected to protect the ApoI site-2, and hence the scaffold is expected to be digested fully at the ApoI site-2. The first lane in Figure 5b corresponds to the 130 bp-1 scaffold, where two bands, corresponding to 113 bp and 106 bp, can be seen. This result is in agreement with our findings above that even a short, 3-nm-long, nucleoprotein filament can self-assemble site-specifically onto a dsDNA scaffold with assembly yields of around 40% and thus protect the ApoI site-1 from digestion. This is in contrast to the results of the restriction digest of scaffolds 130 bp-2 and 130 bp-3, which are shown in lanes 2 and 3 in Figure 5b. Here, only one band, corresponding to 101 bp and 89 bp for the 130 bp-2 and 130 bp-3 scaffolds, respectively, can be seen arising from the full digestion at both ApoI sites. This shows that, in both cases, for the 130 bp-2 scaffold where the center of the ApoI site-1 is 6 bases away from the 5′ end of the nucleoprotein filament, as well as for the 130 bp-3 scaffold where the center of the ApoI site-1 is 18 bases away from the 5′ end of the nucleoprotein filament, the ApoI site-1 was not protected and hence digested. These findings clearly demonstrate the site-specificity of the assembly mechanism, even for a nucleoprotein filament of only 3 nm in length and comprising only two monomers of RecA.
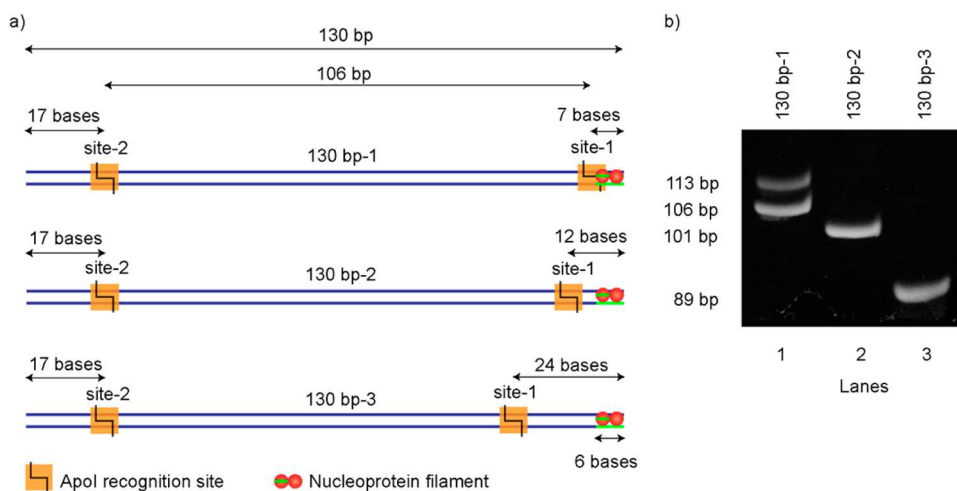
Figure 5. (a) Schematic diagram showing the locations of the ApoI site-1 and site-2 and the binding region of the nucleoprotein filament created from a 6-nucleotide-long ssDNA molecule on the 130 bp-1, 130 bp-2, and 130 bp-3 DNA scaffolds. The positions of ApoI site-2 as well as the binding region of the nucleoprotein filament are the same for all three scaffolds, whereas the ApoI site-1 is centered 7 bases, 12 bases, and 24 bases away from the end in the 130 bp-1, 130 bp-2, and 130 bp-3 scaffold, respectively. (b) 15% PAGE gel (19:1, 80 V, 5 h) demonstrating the sequence-specificity of RecA-mediated assembly on a DNA scaffold using nucleoprotein filaments created from 6-base-long ssDNA molecules, with a length of 3 nm. The nucleoprotein filament partially protects the ApoI site-1 on the 130 bp-1 DNA from ApoI digestion (lane 1); however it does not protect the ApoI site-1 on the 130 bp-2 and 130 bp-3 scaffolds (lanes 2 and 3). The ApoI site-2 is not protected by the nucleoprotein filaments in any of the dsDNA scaffolds.

CONCLUSIONS

We have demonstrated that the protein RecA can be polymerized around single-stranded DNA molecules ranging in length from 60 nucleotides down to 6 nucleotides to form nucleoprotein filaments. If formed in the presence of the nonhydrolyzable ATP analogue ATPγS, these nucleoprotein filaments are stable and can be site-specifically assembled onto double-stranded DNA scaffolds featuring regions that are homologous to the constituent single-stranded DNA used to form the nucleoprotein filament. We have designed two distinct DNA scaffolds, one with a region of homology at the end of the scaffold and one with a region of homology 51 base-pairs away from the end. In both cases, all nucleoprotein filaments investigated, including the shortest filaments created from 6-nucleotide-long single-stranded DNA molecules, could be successfully and site-specifically assembled onto the region of homology on the double-stranded DNA scaffold. The shortest nucleoprotein filaments investigated are 3 nm long, are built from only two monomers of RecA, and cover significantly less than one DNA helical turn when assembled. Finally, we demonstrated that the RecA-driven self-assembly process is highly site-specific, and its precision is only limited by the size of individual DNA bases. A nucleoprotein filament assembled onto a region on the double-stranded DNA scaffold overlapping partly with the recognition site of a restriction enzyme prevents the scaffold from being digested. In contrast, when assembled onto a region on the scaffold that is only 6 bases away from the center of the recognition site, no protection against digestion is observed.

These findings represent a significant improvement over previously established limits of RecA-facilitated self-assembly and position this technique as a valuable nanoscale tool. They open up a range of previously inaccessible avenues for programmable site-specific self-assembly and enable, for example, the synthesis of functional nanostructured materials and, in particular, materials that are heterogeneously structured on the nanoscale.

MATERIALS AND METHODS

Materials. RecA protein (from *E. coli*) at a concentration of 2 mg/mL stored in 20 mM Tris-HCl (pH 7.5), 1 mM DTT, 0.1 mM EDTA, and 50% glycerol was purchased from New England Biolabs (NEB), Inc. (Ipswich, MA, USA). ApoI enzyme at a concentration of 10 units/μL and tango buffer containing 33 mM tris-Acetate (TAc), 10 mM magnesium acetate (MgAc), 66 mM potassium acetate, and 0.1 mg/mL BSA were purchased from Thermo Fisher Scientific Inc. HPLC-purified synthetic oligonucleotides were purchased from Eurofins MWG Operon (Ebersberg, Germany). Proteinase-K from *Tritirachium album* at a concentration of

>800 units/mL; adenosine 5′-[γ-thio]triphosphate tetralithium (ATPγS) salt, which was dissolved in deionized water to a final concentration of 5 mM; 5× Tris-borate-EDTA (TBE) buffer; TAc that was made up to 300 mM pH 7.4); MgAc that was made up to 200 mM pH 7.4; and acrylamide/bis-acrylamide (19:1) 30% solution were all purchased from Sigma Aldrich (St. Louis, MO, USA).

All the dsDNA scaffolds used in this study were generated by PCR amplification from a modified pGEM-T plasmid vector. The nucleotide sequence from which all dsDNA scaffolds were generated and the primer sets used are provided in the Supporting Information.

Methods. The generation of the nucleoprotein filaments and the methods used for their self-assembly onto dsDNA scaffolds were the same as reported previously.[7] Briefly, 1 μL of ssDNA oligonucleotides (60 bases to 6 bases) at 40 μM was mixed with RecA protein (at a concentration of one RecA monomer per three nucleotides) and 1 mM ATPγS in a buffer containing 30 mM TAc and 2 mM MgAc pH 7.4. The sample was incubated for 15 min at 37 °C. To sequester the excess of unbound but functionally active RecA proteins, 1 μL of 30-base-long oligo-(T)$_{30}$ molecules at a concentration of 100 μM was added, and the sample was incubated for an additional 15 min at 37 °C. The above reaction was mixed with 1 μM dsDNA scaffold, in 30 mM TAc and 20 mM MgAc pH 7.4 buffer. The sample was then incubated for 45 min at 37 °C. Ten units of ApoI in 1× tango buffer (final concentration) was added to the above sample, which was incubated for 15 min at 37 °C. The reaction was stopped by adding 5 μL of proteinase K (at concentration >800 units/ml) for 30 min at 37 °C. The samples were then run on a 13−15% polyacrylamide gel (19:1) in 1× TBE buffer for 5−9 h at 80−120 V and were subsequently analyzed with Kodak 1D gel image analysis software. The normalized intensities of all the bands of the gels were determined by Image J software.

For analyzing the patterned DNA scaffolds using AFM, the ratio of the number of nucleoprotein filaments to the dsDNA was kept as 3:1. The pattern formation was allowed to proceed for 45 min at 37 °C without sequestering the excess RecA with the 30-base-long oligo-(T)$_{30}$. The prepared sample was diluted (8−10 ng of the dsDNA in total) in deionized water and was allowed to adsorb on a freshly cleaved mica surface, pretreated with 10 mM NiCl$_2$ for 10 min. The mica surface was washed twice with deionized water and dried thoroughly with a stream of nitrogen. The resulting sample was imaged with a Veeco Dimension 3100 AFM in the tapping mode, in air, with an OTESPA Olympus etched Si cantilever (tetrahedral tip, force constant 42 N/m, resonant frequency around 300 kHz, Al reflective coating).

Conflict of Interest: The authors declare no competing financial interest.

Supporting Information Available: Sequences of all DNA scaffolds used in this work. This material is available free of charge *via* the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

1. Wei, B.; Dai, M.; Yin, P. Complex Shapes Self-Assembled from Single-Stranded DNA Tiles. *Nature* **2012**, *485*, 623–626.
2. Seeman, N. C. Nanomaterials Based on DNA. *Annu. Rev. Biochem.* **2010**, *79*, 65–87.
3. Rothemund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440*, 297–302.
4. Tosch, P.; Walti, C.; Middelberg, A. P. J.; Davies, A. G. Generic Technique to Generate Large Branched DNA Complexes. *Biomacromolecules* **2006**, *7*, 677–681.
5. Liu, J.; Geng, Y.; Pound, E.; Gyawali, S.; Ashton, J. R.; Hickey, J.; Woolley, A. T.; Harb, J. N. Metallization of Branched DNA Origami for Nanoelectronic Circuit Fabrication. *ACS Nano* **2011**, *5*, 2240–2247.
6. Park, S. H.; Barish, R.; Li, H.; Reif, J. H.; Finkelstein, G.; Yan, H.; LaBean, T. H. Three-Helix Bundle DNA Tiles Self-Assemble into 2D Lattice or 1D Templates for Silver Nanowires. *Nano Lett.* **2005**, *5*, 693–696.
7. Sharma, R.; Davies, A. G.; Walti, C. Nanoscale Programmable Sequence-Specific Patterning of DNA Scaffolds Using RecA Protein. *Nanotechnology* **2012**, *23*, 365301.
8. Sharma, R.; Davies, A. G.; Walti, C. RecA Protein Mediated Nano-Scale Patterning of DNA Scaffolds. *J. Nanosci. Nanotechnol.* **2011**, *11*, 10629–10632.
9. Nishinaka, T.; Takano, A.; Doi, Y.; Hashimoto, M.; Nakamura, A.; Matsushita, Y.; Kumaki, J.; Yashima, E. Conductive Metal Nanowires Templated by the Nucleoprotein Filaments, Complex of DNA and RecA Protein. *J. Am. Chem. Soc.* **2005**, *127*, 8120–8125.
10. Keren, K.; Krueger, M.; Gilad, R.; Ben-Yoseph, G.; Sivan, U.; Braun, E. Sequence-Specific Molecular Lithography on Single DNA Molecules. *Science* **2002**, *297*, 72–75.
11. Van der Heijden, T.; Modesti, M.; Hage, S.; Kanaar, R.; Wyman, C.; Dekker, C. Homologous Recombination in Real Time: DNA Strand Exchange by RecA. *Mol. Cell* **2008**, *30*, 530–538.
12. Bazemore, L. R.; Folta-Stogniew, E.; Takahashi, M.; Radding, C. M. RecA Tests Homology at Both Pairing and Strand Exchange. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11863–11868.
13. Cox, M. Motoring along with the Bacterial RecA Protein. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 127–138.
14. Kowalczykowski, S. C. Structural Biology: Snapshots of DNA Repair. *Nature* **2008**, *453*, 463–466.
15. Chen, Z.; Yang, H.; Pavletich, N. Mechanism of Homologous Recombination from the RecA-ssDNA/dsDNA Structures. *Nature* **2008**, *453*, 489–484.
16. Joo, C.; McKinney, S. A.; Nakamura, M.; Rasnik, I.; Myong, S.; Ha, T. Real-Time Observation of RecA Filament Dynamics with Single Monomer Resolution. *Cell* **2006**, *126*, 515–527.
17. Bianco, P. R.; Weinstock, G. M. Interaction of the RecA Protein of Escherichia coli with Single-Stranded Oligodeoxyribonucleotides. *Nucleic Acids Res.* **1996**, *24*, 4933–4939.
18. Leahy, M.; Radding, C. Topography of the Interaction of RecA Protein with Single-Stranded Deoxyoligonucleotides. *J. Biol. Chem.* **1986**, *261*, 6954–6960.
19. Hsieh, P.; Camerini-Otero, C. S.; Camerini-Otero, R. D. The Synapsis Event in the Homologous Pairing of DNAs: RecA Recognizes and Pairs Less than One Helical Repeat of DNA. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6492–6496.
20. Hsieh, P.; Camerini-Otero, C. S.; Camerini-Otero, R. D. Pairing of Homologous DNA Sequences by Proteins: Evidence for Three-Stranded DNA. *Genes Dev.* **1990**, *4*, 1951–1963.
21. Volodin, A. A.; Smirnova, H. A.; Bocharova, T. N. Efficient Interaction of RecA Protein with Fluorescent Dye-Labeled Oligonucleotides. *FEBS Lett.* **1994**, *349*, 65–68.
22. Ragunathan, K.; Joo, C.; Ha, T. Real-Time Observation of Strand Exchange Reaction with High Spatiotemporal Resolution. *Structure* **2011**, *19*, 1064–1073.
23. Cheng, S.; Van Houten, B.; Gamper, H. B.; Sancar, A.; Hearst, J. E. Use of Psoralen-Modified Oligonucleotides to Trap Three-Stranded RecA-DNA Complexes and Repair of These Cross-Linked Complexes by ABC Excinuclease. *J. Biol. Chem.* **1988**, *263*, 15110–15117.
24. Rao, B. J.; Chiu, S. K.; Radding, C. M. Homologous Recognition and Triplex Formation Promoted by RecA Protein between Duplex Oligonucleotides and Single-Stranded DNA. *J. Mol. Biol.* **1993**, *229*, 328–343.
25. Rosselli, W.; Stasiak, A. Energetics of RecA-Mediated Recombination Reactions. Without ATP Hydrolysis RecA Can Mediate Polar Strand Exchange but Is Unable to Recycle. *J. Mol. Biol.* **1990**, *216*, 335–352.
26. VanLoock, M. S.; Yu, X.; Yang, S.; Lai, A. L.; Low, C.; Campbell, M. J.; Egelman, E. H. ATP-Mediated Conformational Changes in the RecA Filament. *Structure* **2003**, *11*, 187–196.
27. Galletto, R.; Amitani, I.; Baskin, R. J.; Kowalczykowski, S. C. Direct Observation of Individual RecA Filaments Assembling on Single DNA Molecules. *Nature* **2006**, *443*, 875–878.
28. van Loenhout, M. T. J.; van der Heijden, T.; Kanaar, R.; Wyman, C.; Dekker, C. Dynamics of RecA Filaments on Single-Stranded DNA. *Nucleic Acids Res.* **2009**, *37*, 4089–4099.
29. Egelman, E. H.; Stasiak, A. Structure of Helical RecA-DNA Complexes: Complexes Formed in the Presence of ATP-Gamma-S or ATP. *J. Mol. Biol.* **1986**, *191*, 677–697.
30. Lusetti, S. L.; Shaw, J. J.; Cox, M. M. Magnesium Ion-Dependent Activation of the RecA Protein Involves the C Terminus. *J. Biol. Chem.* **2003**, *278*, 16381–16388.

31. Bell, C. Structure and Mechanism of Escherichia coli RecA ATPase. *Mol. Microbiol.* **2005**, *58*, 358–366.

32. Forget, A. L.; Kowalczykowski, S. C. Single-Molecule Imaging of DNA Pairing by RecA Reveals a Three-Dimensional Homology Search. *Nature* **2012**, *482*, 423–427.

33. Ragunathan, K.; Liu, C.; Ha, T.; Zhuang, X. RecA Filament Sliding on DNA Facilitates Homology Search. *eLife* **2012**, *1*, e00067.

34. Yancey-Wrona, J. E.; Camerini-Otero, R. D. The Search for DNA Homology Does Not Limit Stable Homologous Pairing Promoted by RecA Protein. *Curr. Biol.* **1995**, *5*, 1149–1158.

35. Shigemori, Y.; Oishi, M. Stable Triple-Stranded DNA Formation and Its Application to the SNP Detection. *DNA Res.* **2005**, *12*, 441–449.