Protein Nanoarrays

Aptamer-Directed Self-Assembly of Protein Arrays on a DNA Nanostructure**

*Yan Liu, Chenxiang Lin, Hanying Li, and Hao Yan**

DNA-based self-assembly[1] represents a versatile system for nanoscale construction due largely to the well-characterized conformation of DNA and its predictability in the formation of base pairs. The methodology of DNA self-assembly begins with the chemical synthesis of single-stranded DNA molecules that self-assemble into branched DNA motifs, known as tiles. DNA tiles can carry sticky ends that are complementary to the sticky ends of other DNA tiles; this facilitates continued assembly of the tiles into DNA lattices. Substantial progress has been made in recent years with DNA nanostructures in the construction of patterned arrays.[2–13] One promising application of DNA-based self-assembly is the use of self-assembled DNA nanostructures to direct the assembly of other macromolecular components such as proteins. DNA-templated protein arrays with predictable control at the nanometer scale could lead to single-molecule detection in proteomics studies. Individual proteins placed at unique locations on the nanoarray could be detected with single-molecule imaging techniques such as recognition imaging, in which specific antibodies are attached to the scanning probe cantilever.[14] Niemeyer has also pointed out another interesting application of precisely controlled protein assemblies in the construction of spatially well-defined multienzyme constructs.[15]

The noncovalent streptavidin–biotin interaction has been the most-explored method in the creation of streptavidin molecule networks through the modification of DNA oligonucleotides with biotin.[6,16–19] However, this method is limited in that only one type of protein–ligand interaction is present. Although other proteins can be fused with streptavidin by protein engineering, it is a tedious process, and the function of the proteins and the biotin-binding capacity of streptavidin may both be affected by the fusion. The covalent linkage of proteins to nucleic acids often relies on the use of heterobispecific cross-linkers,[20] and involves multistep chemical coupling and purification steps. Another method to covalently link proteins to oligonucleotides uses recombinant protein engineering to incorporate a cysteine residue into the protein to allow subsequent coupling to thiolated oligonucleotides by disulfide linkages.[21–23] Although attachment through covalent bonds would be ideal, the process for this has not yet been automated. To realize the potential of self-assembled DNA templates in the construction of protein nanoarrays, there is an immediate need to develop programmable methods in which DNA nanostructures are used to direct the assembly of any protein of interest.

Aptamers are DNA or RNA molecules that can be selected from random pools based on their ability to bind other molecules. Aptamers have been selected to bind other nucleic acids, proteins, small organic compounds, and even entire organisms.[24–27] Aptamers that exhibit sub-nanomolar affinities for a wide range of protein targets have been identified.[27,28] It is possible to generate a virtually unlimited number of specific ligand–aptamer pairs, so that each class of spatially displayed aptamer will interact with high affinity to its specific ligand. Cox and Ellington have reported the identification of new aptamer sequences through an automated systematic evolution of ligands by exponential enrichment (SELEX) process.[25]

Herein, we demonstrate the first use of selective DNA aptamer binding as a robust platform to link proteins to periodic sites of a self-assembled DNA array. The system employs three components: 1) a rationally designed DNA nanostructure that self-assembles into highly ordered spatial lattices by virtue of specific annealing of complementary sticky ends; 2) a DNA-docking site containing an aptamer sequence which tethers the protein of interest to the DNA lattice; and 3) the protein to be displayed on a self-assembled DNA lattice.

The DNA aptamer-directed self-assembled protein nanoarray possesses the following advantageous features: 1) the DNA tiling self-assembly is programmable: a rich set of DNA tiles and lattices with various geometries and patterns can be generated by altering the tiling design and sticky-end associations;[2–13] 2) the DNA tiles and aptamers are compatible with each other, as they are both composed of oligonucleotides; 3) new protein-binding aptamers can be generated with automated processes; 4) DNA nanostructure-displayed aptamers are not just limited to proteins, but can be extended to bind other ligand types.

To demonstrate this system, we chose the thrombin-binding aptamer (TBA). TBA is a well-characterized 15-base DNA aptamer with a consensus sequence of d(GGTTGGTGTGGTTGG) that folds into a unimolecular guanine quadruplex and binds thrombin with nanomolar affinity.[29] Thrombin is a multifunctional serine protease that

[*] Dr. Y. Liu, C. Lin, H. Li, Prof. Dr. H. Yan
Department of Chemistry and
Biochemistry & Biodesign Institute
Arizona State University
Tempe, AZ 85287 (USA)
Fax: (+1) 480-965-2747
E-mail: hao.yan@asu.edu

[**] This work was supported by grants from NSF (CCF-0453686, CCF-0453685) and a research grant from the Biodesign Institute at ASU to H.Y. We thank Prof. Andrew Ellington, Prof. Thomas H. LaBean, and Prof. Daniel Kenan for helpful discussions.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

 DOI: 10.1002/ange.200501089

recognizes multiple macromolecular substrates and plays a key role in both coagulation and anticoagulation.

Figure 1 illustrates the use of a triple-crossover (TX) DNA tile as the template to direct the assembly of an aptamer and its subsequent organization of proteins into 1D periodic arrays. The TX tile shown is similar to another developed previously,[3,19] except that it has two DNA hairpin loops: one protrudes outward in the plane of the tile and contains the

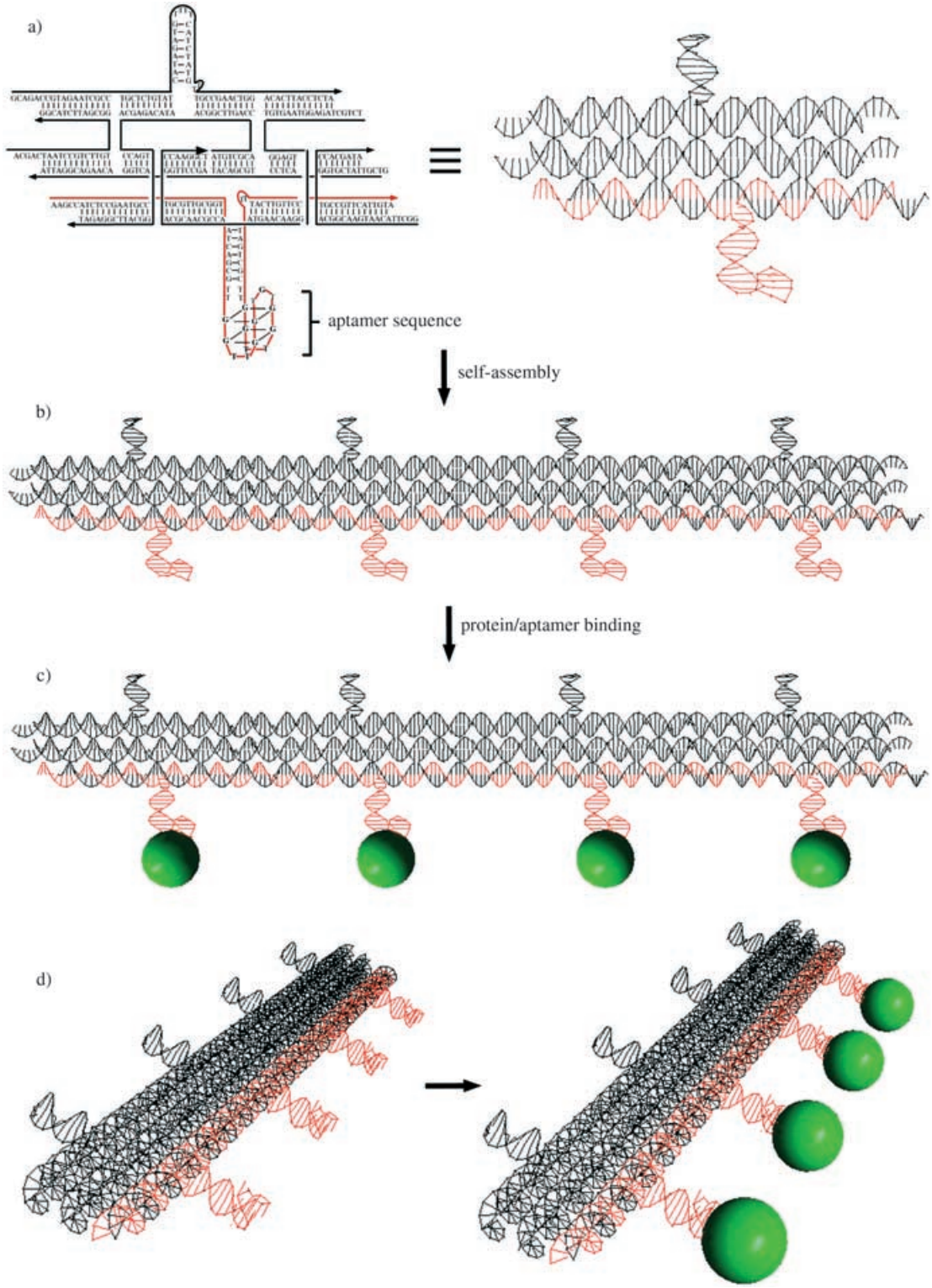

Figure 1. Aptamer-directed self-assembly of thrombin protein on a triple crossover DNA tile: a) DNA strand structure and sequence used for the self-assembly, the red strand contains the thrombin-binding aptamer sequence, which is illustrated as a G-quadruplex structure; b) the TX tiles self-assemble into linear DNA nanoarrays, in which the spacing between adjacent aptamer loops corresponds to five helical DNA turns: ≈17 nm; c) binding of thrombin (green spheres) to aptamers on the linear array leads to linear protein arrays; d) 3D rendering of the aptamer-directed self-assembly of the protein array.

aptamer sequence (in red) at the end of the stem, the other loop serves as a control; it also protrudes out of the TX tile, but does not contain the aptamer sequence. The length of the stem is arbitrary and can be adjusted as needed to place the target binding molecules in the desired positions and rotational orientations. In general, aptamers can be extended from the tile lattice regardless of their specific secondary structures. Furthermore, the DNA lattice may contain an arbitrary number of different tiles if necessary, so that the aptamers that bind proteins can be placed at every tile, every other tile, every third tile, and so on. Notably, various protein-binding tiles within the lattice may potentially be engineered to display different proteins at defined positions and orientations by virtue of the different possible combinations of aptamers and target ligands. Together, the DNA nanostructure and aptamers constitute a versatile and novel platform for the nanoarchitecture of proteins and synthetic molecules.

The dimensions of the three components used in this study are summarized below. The TX tile is approximately $6 \times 17 \times 2$ nm^3 as illustrated in Figure 1 a. The DNA aptamer stem loop is approximately 2 nm in helical diameter. A stem of 9 bp ($l \approx 3$ nm) was included to allow enough space for the thrombin to bind the aptamer. Thrombin has a molecular weight of ≈ 37 kDa, and a spherical diameter of about 3 nm.[30] Other larger target macromolecules could also be accommodated onto the lattice by varying the dimension of the tiles and the length of the stems. The TX tile self-assembles in $1 \times$ TAE/Mg^{2+} buffer (Experimental Section) into a linear array of TBA units with a periodic distance of ≈ 17 nm between two adjacent aptamers (Figure 1 b). The self-assembly of the DNA linear array is then followed by the addition of thrombin protein to the solution. Binding of thrombin to its aptamer results in a periodic linear array of thrombin molecules, illustrated by the green spheres in Figure 1 c. Figure 1 d gives a 3D view of the aptamer-directed self-assembly of thrombin protein on the 1D TX array.

Both AFM and gel electrophoresis experiments were used to demonstrate aptamer-directed self-assembly. Following formation of the 1D TX array and subsequent binding of thrombin, the preparations were examined by AFM. Figure 2 a shows an AFM height profile of the linear TX arrays before the addition of thrombin protein. Figures 2 b–d show the protein nanoarrays that result from the binding of thrombin to the periodic aptamer sites on the linear TX array. With its diameter of ≈ 3 nm, thrombin binding to the aptamer sequences generates topographical features on the mica surface that are higher than those of bare TX arrays; these are observed as regularly spaced brighter spots at the aptamer locations. The AFM images of Figures 2 b–d clearly demonstrate the regular spacing of the thrombin molecules templated on the linear TX arrays. The average distance measured between pairs of adjacent thrombin molecules along a given TX array is $\approx 18 \pm 1$ nm (Figure 3, profile 1), which is in good agreement with the designed parameters of five helical DNA turns between each adjacent aptamer pair. AFM measurements show that thrombin molecules have an average height of $\approx 2.7 \pm 0.3$ nm (Figure 3, profiles 3–6), in comparison with a height of $\approx 1.7 \pm 0.2$ nm measured across

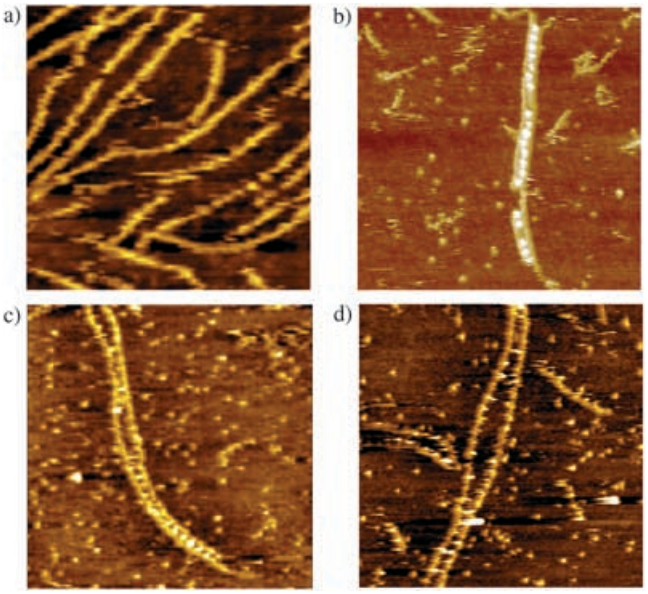

Figure 2. AFM images of the aptamer-directed self-assembly of thrombin protein linear arrays: a) 1D TX tile array with aptamer sequences prior to thrombin protein binding; b)–d) AFM images after binding of thrombin to the TX array; brighter spots correspond to thrombin proteins. Each image shows 500×500 nm^2.

the DNA segment on the array (Figure 3, profile 2). This further confirms that the periodic bumps in the lattice result from the binding of thrombin to the DNA tiles. (More distance and height profiles are available in the Supporting Information.)

Interestingly, upon binding with thrombin, the TX linear array has an inclination to form a parallel pair with the proteins sandwiched inside. One possible explanation is that each protein binds with two aptamers which brings two opposing TX tiles together. A second possibility is that the protein dimerizes under the experimental conditions, and each monomer in the dimer binds with an aptamer on the opposite side of the TX tiles, thus creating the parallel pair of the TX linear array. We believe the latter explanation is most relevant, and this conclusion is supported by several important observations: first, the average center-to-center distance between the parallel pair of DNA linear arrays is 22 ± 1 nm (Figure 3, profile 2). The width of the TX tile is ≈ 6 nm, the hairpin loop including the aptamer sequence on each TX tile is about 4–5 nm, and therefore the protein structure in the center should have a span of ≈ 7–8 nm. This spacing is consistent with the hypothesis of two protein molecules separating the DNA arrays. Secondly, some loosely bound TX pairs were also observed with alternating, closely parallel, separated sections (Figures 2 c and d). Upon careful examination of the separated parts, proteins were observed to be bound on one side but not on the other. This indicates that the linkage between the two parallel linear DNA arrays is not through a single protein molecule, but through a protein dimer. Upon examination of the X-ray crystal structure of the thrombin–aptamer complex, each thrombin molecule is observed to have two possible binding sites for the different locations of the aptamer.[30] Specifically, the G8-T9-G10

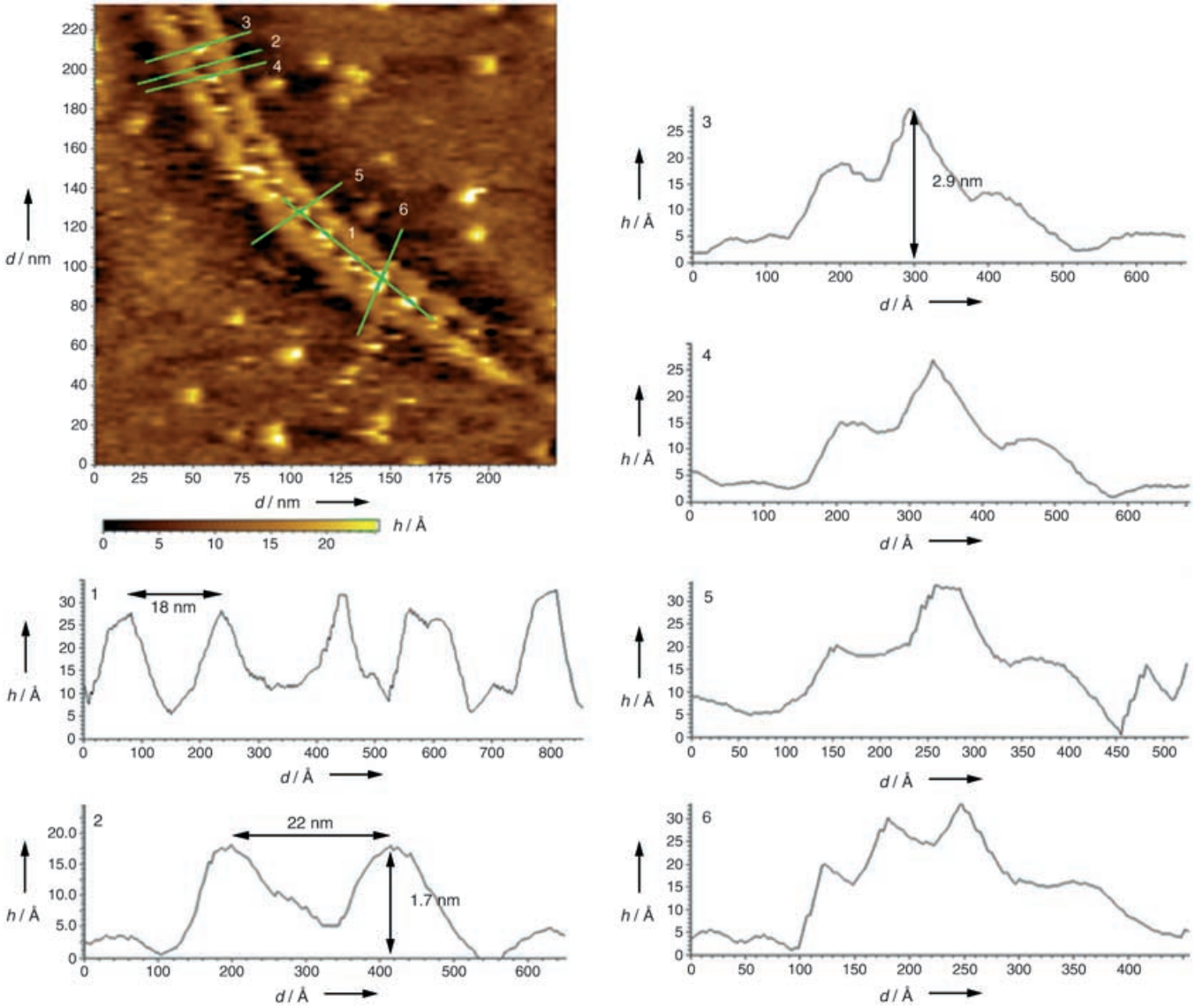

Figure 3. Analysis of the AFM data in Figure 2c (magnified): lines and numbers in the image correspond to the cross sections of the image shown in profiles 1 to 6. The periodicity of the protein molecules in the array is clearly shown (lateral distance ≈17–19 nm). The parallel pairing of the TX linear arrays show protein molecules sandwiched in between that are higher (2.5–3 nm), than the DNA tiles (≈1.7 nm). Profile 6 clearly shows two loosely bound thrombin molecules resolved by AFM. The average distance between the linear DNA arrays is ≈22 nm.

region of the aptamer binds the fibrinogen-binding site (Arg75 and Arg77), and the T13-G14-G15 region binds to the heparin-binding site (Arg101, Arg233, Lys236, Trp237, and Lys240) through both ion pairing (negatively charged phosphate on the DNA backbone with the positively charged amino acid residues on the protein) and hydrogen bonding.[30] In the crystal structure of the complex, the aptamer is squeezed in between two thrombin molecules and concurrently, one thrombin protein has two aptamer-binding sites. Although the site of primary bimolecular interaction in solution is not clear, we propose that both sites in both the aptamer and protein molecules are involved, and the dimerization of thrombin is induced by aptamer binding. Owing to the limit of AFM resolution, the geometry of dimer formation is not clear. It is possible that the two opposing TX arrays can assume parallel or antiparallel alignments.

The model proposed above is further evidenced by analysis with nondenaturing polyacrylamide gel electrophoresis. Figure 4 shows a gel image that demonstrates the concentration dependence of TX aptamer–protein binding. The sticky ends of the TX tiles were removed to allow the formation of single tile complex monomers, which are able to migrate in the gel. The isoelectric point (pI) of thrombin is 7.5, close to the pH value of the buffer. Therefore, the protein alone barely migrates in the gel. However, if bound to the negatively charged TX tile, the thrombin–aptamer complex can migrate in the gel, albeit slower than TX tile alone. Stains-all was used for gel imaging; as the protein is neutral in charge, it is not stained with the positively charged dye, whereas the DNA is well-stained.

The gel image shows that with an increase in protein concentration, at least two slower-migrating bands show up in

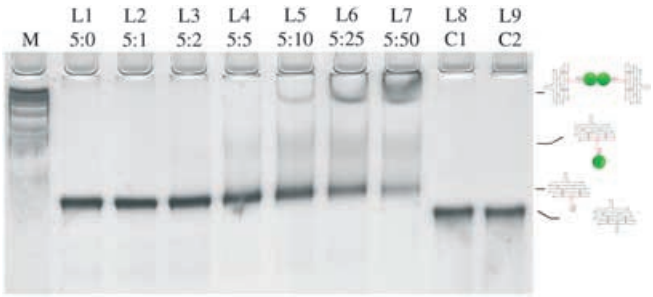

Figure 4. Nondenaturing gel electrophoresis of TX tile and TX aptamer–protein complexes. The polyacrylamide gel (10%) in 1×TAE/Mg^{2+} buffer was run at constant current (25 mA) for 48 h at 4°C. The gel was stained with Stains-all (0.1%) and destained by exposure to white light. Lane M: 100-bp DNA ladder; lane L1: TX tile alone; lanes L2–L7: consistent quantity of TX tiles (10 pmol) but an increasing amount of thrombin from 2 to 100 pmol. DNA/protein ratios are indicated for each lane. Lane L8 contains the control TX molecule without the aptamer sequence; L9 contains a mixture of the control TX and thrombin at a molar ratio of 5:10. The assignments of the bands are indicated schematically on the right-hand side.

the lane, and the intensity distribution of the bands within the lane shifts toward the slower-migrating bands. At a molar ratio of 5:1 TX tile/thrombin, the first slower-migrating band initially appears, the intensity of which increases and then levels off at a tile/protein ratio of 5:10. At a molar ratio of about 5:10, another even slower-migrating band appears; its intensity increases steadily up to the ratio of 5:50. If the aptamer–protein interaction is a one-to-one stoichiometry, only one slower-migrating band would be expected to appear. The appearance of the second slower band and its relative position in the gel indicate that this is a dimer of the TX DNA molecule with one or two protein molecules sandwiched in between. If only one protein molecule were present between each TX pair, the intensity of this band relative to the first slower-migrating band would be expected to decrease with an increase in protein concentration, yet the opposite trend is observed. We therefore conclude that two protein molecules are sandwiched between two opposing TX tiles, as indicated by the schematic representation to the right of the gel image. Lanes 8 and 9 are the controls: lane 8 contains the TX molecule without the aptamer sequence (strand structure and sequence shown in Supporting Information); it migrates slightly faster than the TX assembly with the aptamer sequence in the lane 2. Lane 9 contains a 5:10 molar ratio of the control TX tile used in lane 8 and thrombin. No slower-migrating bands are evident, which indicates that nonspecific interactions are not present between the protein and the TX tile in which the aptamer is absent. It is also notable that the fast-migrating bands show a backward shift in the lanes with the high protein concentrations, whereas the middle bands smear forward. This can be explained through an unstable interaction between the TX aptamer and the thrombin molecule in the 1:1 complex; such complexes formed in solution may break down under gel conditions. To obtain good band separation, the polyacrylamide gel (10%) was run for 48 h. The thrombin–aptamer interaction is governed not by covalent bonds, but by dynamic equilibrium. It is possible that during the long gel-running time, some complex dissociation occurs. This would result in the DNA aptamer band running in the front, but hindered, as observed, and the 1:1 protein–aptamer complex running in the middle, yet smeared.

On the other hand, the 2:2 complex (the slowest-migrating band) is apparently more stable. This phenomenon is consistent with the observation that at a molar ratio of 5:10, AFM images show that aptamers on the DNA array are mostly occupied by thrombin. Overall, dimer formation as evidenced by gel electrophoresis is consistent with the AFM imaging results.

In summary, we have demonstrated the incorporation of aptamer sequences into a rationally designed DNA nanostructure, and have successfully used the aptamer-bearing DNA nanostructure for the directed assembly of thrombin protein arrays. The results clearly show that the thrombin-binding aptamer still functions as the protein-binding moiety upon incorporation into a complex DNA nanostructure. This is encouraging, as it will open up future opportunities for the construction of nanoscale protein arrays in a programmable fashion.

The DNA tiles could be easily modified to accommodate other aptamer sequences in the construction of 2D or 3D protein arrays. Precisely controlled organization of protein molecules onto periodic 2D DNA lattices would give insight into protein structure with 2D cryoelectron diffraction microscopy. However, the ultimate resolution for obtaining structural information from spatially organized proteins will depend on the flexibility of the orientation, and the conformation of the aptamers on the DNA tiles. A recent development by Simmel and co-workers[31] involved a combination of the aptamer-directed self-assembly system with the use of fuel DNA strands to control the binding and release of thrombin proteins. In this way, the pattern of the protein nanoarray could be efficiently tuned. Also, the incorporation of previously developed DNA nanoactuators[32] into the aptamer-directed protein array would permit a change in the relative positions of proteins in real time, which could enable the study of proximity effects of protein–protein interactions. Indeed, the DNA-directed self-assembly of proteins can be used for the direct visualization of protein–protein interactions, as the spatial resolution of such interactions is greatly amplified by the patterning of the DNA structure.

Experimental Section

All DNA strands used for the work reported herein are listed in the Supporting Information. The DNA sequences were designed with the SEQUIN program.[33] DNA strands were commercially synthesized (Integrated DNA Technologies, Inc.) and purified by denaturing gel electrophoresis. Thrombin from human plasma (freeze-dried from a sodium citrate buffer) and Stains-all were purchased from Sigma-Aldrich. The TX tile complex and DNA array were formed by mixing equal quantities of each strand designed in the complex or lattice unit at a concentration of 1 μM (as estimated by OD_{260}) in 1×TAE/Mg^{2+} buffer (Tris, 40 mM; acetic acid, 20 mM; EDTA, 2 mM; and magnesium acetate, 12.5 mM; pH 8.0). The mixture was cooled slowly from 90°C to 20°C. Thrombin was reconstituted into aqueous solution by adding ultrapure water to a final protein concentration of ≈10 μM.

Nondenaturing gel electrophoresis: TX tiles were mixed with thrombin at different concentration ratios and left at room temperature for 30 min to establish binding equilibrium. Sampling aliquots were loaded onto a nondenaturing polyacrylamide gel (10%), and run at 4°C for 48 h at a constant current of 25 mA. A solution of Stains-all (0.1%) in water/formamide (45:55, v/v) was used to stain the gel. The gel was destained by exposure to white light, and the gel image was collected with a scanner.

AFM imaging: DNA lattice samples (5 μL, ≈30 nM) were dropped onto freshly cleaved mica (Ted Pella, Inc.) and left to adsorb to the surface for 3 min. Buffer (1 × TAE/Mg^{2+}) in volumes of either 30 μL or 400 μL was then added to the drops on the mica. Imaging was performed in a fluid cell in tapping mode on a Multimode NanoScope IIIa (Digital Instruments) or a Pico-Plus AFM (Molecular Imaging) with NP-S tips (Veeco, Inc.). The 3D structure of the TX tile was drawn with Strata 3D software.

Received: March 26, 2005
Published online: June 9, 2005

Keywords: DNA structures · nanostructures · protein arrays · scanning probe microscopy · self-assembly

[1] N. C. Seeman, *Nature* **2003**, *421*, 427.
[2] E. Winfree, F. Liu, L. A. Wenzler, N. C. Seeman, *Nature* **1998**, *394*, 539.
[3] T. H. Labean, H. Yan, J. Kopatsch, F. R. Liu, E. Winfree, J. H. Reif, N. C. Seeman, *J. Am. Chem. Soc.* **2000**, *122*, 1848.
[4] C. D. Mao, W. Q. Sun, N. C. Seeman, *J. Am. Chem. Soc.* **1999**, *121*, 5437.
[5] R. Sha, F. Liu, D. P. Millar, N. C. Seeman, *Chem. Biol.* **2000**, *7*, 743.
[6] H. Yan, S. H. Park, G. Finkelstein, J. H. Reif, T. H. LaBean, *Science* **2003**, *301*, 1882.
[7] H. Yan, T. H. LaBean, L. P. Feng, J. H. Reif, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8103.
[8] P. W. K. Rothemund, N. Papadakis, E. Winfree, *PLoS Biology* **2004**, *2*, 2041.
[9] P. W. K. Rothemund, A. Ekani-Nkodo, N. Papadakis, A. Kumar, D. K. Fygenson, E. Winfree, *J. Am. Chem. Soc.* **2004**, *126*, 16344.
[10] B. Q. Ding, R. J. Sha, N. C. Seeman, *J. Am. Chem. Soc.* **2004**, *126*, 10230.
[11] J. C. Mitchell, J. R. Harris, J. Malo, J. Bath, A. J. Turberfield, *J. Am. Chem. Soc.* **2004**, *126*, 16342.
[12] D. Liu, S. H. Park, J. H. Reif, T. H. LaBean, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 717.
[13] N. Chelyapov, Y. Brun, M. Gopalkrishnan, D. Reishus, B. Shaw, L. Adleman, *J. Am. Chem. Soc.* **2004**, *126*, 13924.
[14] C. Stroh, H. Wang, R. Bash, B. Ashcroft, J. Nelson, H. Gruber, D. Lohr, S. M. Lindsay, P. Hinterdorfer, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12503.
[15] C. M. Niemeyer in *Nanobiotechnology: Concepts, Applications and Perspectives*, (Eds.: C. M. Niemeyer, C. A. Mirkin), Wiley-VCH, Weinheim, **2004**, 227.
[16] C. M. Niemeyer, M. Adler, S. Gao, L. Chi, *Bioconjugate Chem.* **2001**, *12*, 364.
[17] C. M. Niemeyer, M. Adler, S. Gao, L. Chi, *J. Biomol. Struct. Dyn.* **2002**, *20*, 223.
[18] C. M. Niemeyer, M. Adler, B. Pignataro, S. Lenhert, S. Gao, L. Chi, H. Fuchs, D. Blohm, *Nucleic Acids Res.* **1999**, *27*, 4553.
[19] H. Y. Li, S. H. Park, J. H. Reif, T. H. LaBean, H. Yan, *J. Am. Chem. Soc.* **2004**, *126*, 418.
[20] C. M. Niemeyer, T. Sano, C. L. Smith, C. R. Cantor, *Nucleic Acids Res.* **1994**, *22*, 5530.
[21] D. R. Corey, P. G. Schultz, *Science* **1987**, *238*, 1401.
[22] S. Howorka, S. Cheley, H. Bayley, *Nat. Biotechnol.* **2001**, *19*, 636.
[23] R. B. Fong, Z. L. Ding, C. J. Long, A. S. Hoffman, P. S. Stayton, *Bioconjugate Chem.* **1999**, *10*, 720.
[24] E. N. Brody, M. C. Willis, J. D. Smith, S. Jayasena, D. Zichi, L. Gold, *Mol. Diagn.* **1999**, *4*, 381.
[25] J. C. Cox, A. D. Ellington, *Bioorg. Med. Chem.* **2001**, *9*, 2525.
[26] R. C. Conrad, L. Giver, Y. Tian, A. D. Ellington, *Comb. Chem.* **1996**, *267*, 336.
[27] W. Xu, A. D. Ellington, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 7475.
[28] D. E. Tsai, D. J. Kenan, J. D. Keene, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 8864.
[29] R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe, J. Feigon, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 3745.
[30] K. Padmanabhan, K. P. Padmanabhan, J. D. Ferrara, J. E. Sadler, a. Tulinsky, *J. Biol. Chem.* **1993**, *268*, 17651.
[31] W. U. Dittmer, A. Reuter, F. C. Simmel, *Angew. Chem.* **2004**, *116*, 3634; *Angew. Chem. Int. Ed.* **2004**, *43*, 3550.
[32] L. P. Feng, S. H. Park, J. H. Reif, H. Yan, *Angew. Chem.* **2003**, *115*, 4478; *Angew. Chem. Int. Ed.* **2003**, *42*, 4342.
[33] N. C. Seeman, *J. Biomol. Struct. Dyn.* **1990**, *8*, 573.

