Protein–DNA Conjugates

DOI: 10.1002/anie.200701942

A Universal Method for the Preparation of Covalent Protein–DNA Conjugates for Use in Creating Protein Nanostructures**

*Benjamin P. Duckworth, Yan Chen, James W. Wollack, Yuk Sham, Joachim D. Mueller, T. Andrew Taton, and Mark D. Distefano**

DNA has numerous attractive features as a scaffold for nanostructure assembly. Its rigidity, predictable structure, and assembly through complementary hybridization allow DNA to form nanoscale architectures such as cubes,[1] tetrahedra,[2] octahedra,[3,4] and 2D arrays.[5–8] By introducing proteins into DNA nanostructures, the recognition elements and functionalities that are inherent in proteins can be organized into nanostructured motifs. DNA-scaffolded protein assemblies have been used in immuno-PCR detection methods (PCR = polymerase chain reaction)[9–11] to arrange biocatalysts in a series for sequential reactions[12,13] and to organize other nanomaterials.[14] There are currently several methodologies used to link proteins to DNA. Proteins have been assembled onto DNA scaffolds through intervening adapter molecules such as streptavidin[12,15–20] or aptamers.[21,22] Alternately, direct covalent conjugation can be achieved by modification of cysteine or lysine residues[23–26] or intein modification.[11,27,28] Niemyer and co-workers have employed these protein–DNA conjugates to form fluorescence resonant energy transfer (FRET) systems for use in nanobiotechnology.[29,30]

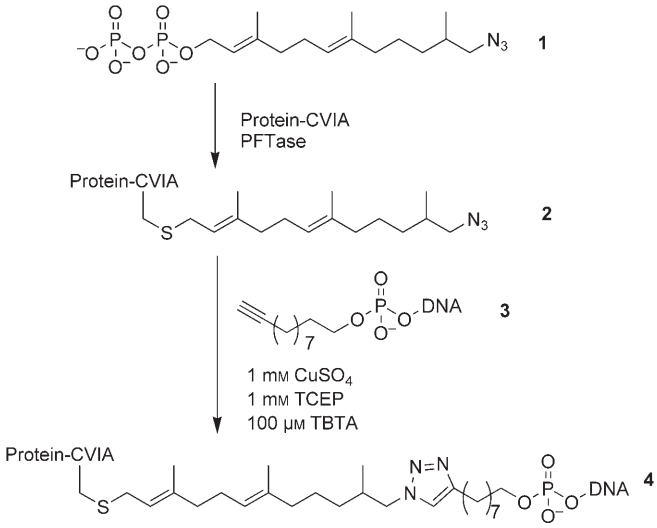
Herein, we demonstrate a fusion-based strategy to regioselectively and covalently label proteins at the C terminus with single-stranded DNA. These protein–oligonucleotide chimeras were then spontaneously assembled into nanoarchitectures by complementary hybridization of the DNA. The covalent attachment strategy described herein yields a short and compact linkage between the protein and DNA molecule that allows for precise control over protein spacing and orientation in the final nanostructure.

[*] B. P. Duckworth, J. W. Wollack, Prof. T. A. Taton, Prof. M. D. Distefano
Department of Chemistry, University of Minnesota
Minneapolis, MN 55455 (USA)
Fax: (+1) 612-626-7541
E-mail: distefan@chem.umn.edu
Homepage: http://www.chem.umn.edu/groups/distafano/
Dr. Y. Chen, Prof. J. D. Mueller
Department of Physics, University of Minnesota
Minneapolis, MN 55455 (USA)
Dr. Y. Sham
Minnesota Supercomputing Institute, University of Minnesota
Minneapolis, MN 55455 (USA)

[**] This work was supported by grants from NIH (GM 008700 and GM 58442). We thank Prof. Andrew Turberfield for the coordinates of the DNA tetrahedron.

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

To achieve selective protein labeling, we use the enzyme protein farnesyltransferase (PFTase) to label a substrate protein containing a C-terminal tetrapeptide tag with an azide-modified isoprenoid diphosphate (**1**, Scheme 1).[31–34]



Scheme 1. PFTase-mediated labeling of target proteins with an azide molecule, and subsequent reaction with alkyne-modified DNA. TBTA = tris(benzyltriazolylmethyl)amine, TCEP = tris(2-carboxyethyl)phosphine.

We have previously shown that this prenylation reaction is complete within 1 h and does not require the target protein or the transferase to be present in their pure forms.[34] Following enzymatic labeling, the azide-functionalized protein (**2**) can be reacted with DNA containing an alkyne (**3**) on its 5′ end by using the Cu^{I}-catalyzed [3+2] Huisgen cycloaddition ("click") reaction. The alkyne-containing phosphoramidite that was used to synthesize **3** was designed with a long carbon chain to aid in the separation of **3** from DNA strands that do not contain the alkyne moiety, however, this feature is not essential.

To illustrate the power of this chemoenzymatic labeling method, we chose enhanced green fluorescent protein (GFP) as a model protein. GFP, expressed and purified to contain a C-terminal CVIA tag, was incubated with **1** and PFTase for 3 h at 30 °C to yield azide-functionalized GFP (**2**). To create alkyne-functionalized DNA, an oligonucleotide that was 33 bases in length was synthesized by using standard solid-phase DNA synthesis techniques. A phosphoramidite containing the hydrophobic carbon chain and an alkyne moiety was manually coupled to the DNA attached to controlled pore

glass (CPG) resin (see the Supporting Information). Deprotection, cleavage, and HPLC purification yielded pure **3**. The click reaction was then used to selectively couple **3** to **2**. The GFP–DNA conjugate (**4**) was purified from unreacted azide-functionalized GFP (**2**) and DNA–alkyne (**3**) material by using gel filtration chromatography (see the Supporting Information). The elution volume of the desired product was easily determined owing to its combination of GFP fluorescence and DNA absorbance, and the efficiency of the conjugation was estimated to be 72%.

To confirm successful conjugation and recognition properties of the modified single-stranded DNA, a complementary 33-mer DNA strand containing a Texas Red fluorophore (TR–DNA) was mixed with the GFP–DNA conjugate (see the Supporting Information). As can be seen in the Supporting Information, when no TR–DNA is present, only GFP fluorescence is observed. When one equivalent of TR–DNA is added, the complementary DNA strands bind, positioning the TR fluorophore close to the GFP so that FRET is observed. A titration of TR–DNA with GFP–DNA is shown in the Supporting Information. The fluorescence of both GFP and TR reach limiting values when the two are present in nearly equivalent concentrations. When TR–DNA was added to GFP molecules that are not conjugated with DNA, no FRET was observed (data not shown). These experiments demonstrate that DNA that has been covalently attached to GFP by using the click reaction behaves as conventional DNA and can hybridize to its complementary strand. To demonstrate the scope of this method, protein–DNA conjugates with CDC42 (a naturally prenylated protein) and glutathione S-transferase were also prepared (see the Supporting Information).

The attachment chemistry described herein can also be employed to construct novel protein–DNA nanoassemblies. As the DNA scaffold onto which proteins can be patterned, we chose the DNA tetrahedron that was recently described by Turberfield and co-workers.[2,35] The tetrahedron consists of four oligonucleotides (**6**, **7**, **8**, and **9**) that self-assemble rapidly to form the superstructure in high yield. Turberfield and co-workers successfully encapsulated a protein molecule within the tetrahedron by conjugating one of the four oligonucleotides to a surface amine of cytochrome *c*.[36]

By using Turberfield's assembly as a model, we designed a tetrahedron with nicks at four of the six edges. Then, component oligonucleotides **10a**, **11a**, **12a**, and **13a** were synthesized by using standard DNA coupling chemistry. To obtain sequence **10a**, three bases from the 5′ end of sequence **6** were transferred to the 3′ end. This shift places the nick so that the oligonucleotide ends would project out of the tetrahedron. Sequences **11a**, **12a**, and **13a** were designed in the same manner from sequences **7**, **8**, and **9**, respectively. A model of the DNA tetrahedron is shown in the left-hand panel of Figure 1b (also see the Supporting Information). The DNA strands were converted to alkyne-functionalized DNA by manual coupling of the alkyne phosphoramidite to the 5′ termini of all four strands to produce **10b**, **11b**, **12b**, and **13b** (see the Supporting Information). Following purification of the alkyne oligonucleotides, GFP–N_3 was individually reacted with each DNA strand by using the click reaction to

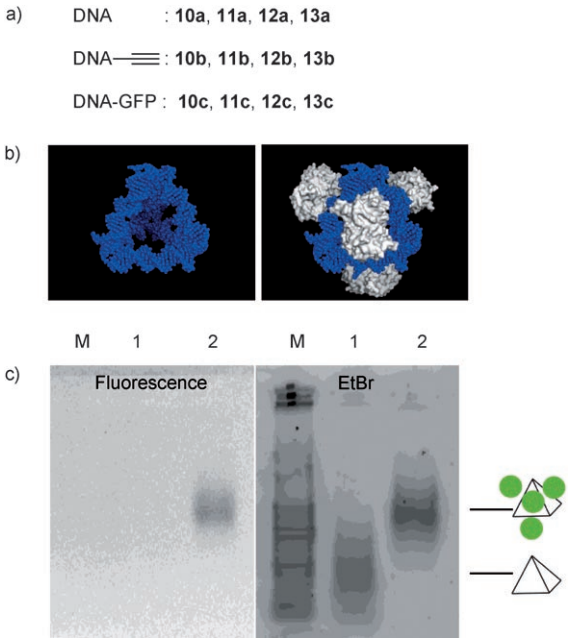


Figure 1. Formation of a protein-modified DNA tetrahedron. a) Numbering system for DNA and GFP–DNA species. b) Molecular modeling of DNA tetrahedron (left) and GFP–DNA tetrahedron (right). DNA is shown in blue and GFP is shown in white. Models were created and energy minimized by using Maestro (Schrödinger). c) Gel analysis of the DNA tetrahedron (lane 1) and the GFP tetrahedron (lane 2). Lane M contains a 200-bp DNA marker.

generate **10c**, **11c**, **12c**, and **13c**. The four GFP–DNA conjugates were then purified by using gel filtration chromatography (see the Supporting Information); analysis of the chromatograms indicated a coupling efficiency between DNA and GFP–N_3 of 70%, which is similar to that observed in the preparation of **4**. To form the GFP tetrahedron (Figure 1b, right-hand panel), the pure GFP–DNA conjugates (**10c**, **11c**, **12c**, and **13c**) were combined so that the final concentration of each conjugate was 1.0 μM in Tris-HCl buffer solution (Tris = tris(hydroxymethyl)aminomethane) containing 5.0 mM $MgCl_2$ (TM buffer solution). As a control, the four oligonucleotides that lacked both the alkyne phosphoramidite and a GFP molecule (**6**, **7**, **8**, and **9**) were also combined. The samples (20 μL) were heated to 54°C for 3 min followed by cooling to 4°C over 2 min. The samples were analyzed by agarose gel electrophoresis at 4°C. The gel was imaged for both green fluorescence and DNA (through staining with ethidium bromide; EtBr). Lane 1 (Figure 1c, right-hand panel) shows a DNA-containing species whose mobility is similar to a 200-bp DNA marker (lane M). This result is consistent with observations reported by Turberfield and co-workers for the formation of the original nonfunctionalized DNA tetrahedron. Lane 2 (Figure 1c, right-hand panel), which contains the GFP-modified DNA strands, shows a DNA-containing species whose mobility is lower than that of the unmodified tetrahedron. As the GFP tetrahedron was designed to position the four GFP moieties outside of the tetrahedron, it was expected that the hydrodynamic radius would be larger and hence the electrophoretic mobility would be lower. Additionally, the species of lower mobility exhibits

green fluorescence when excited at 488 nm (Figure 1c, left-hand panel, lane 2), indicating that it contains both GFP and DNA. Thus, the species observed in lane 2 is consistent with the proposed tetrahedral structure.

To further validate that the GFP–oligonucleotide conjugates self-assemble into the proposed superstructure, tetrahedra with varying numbers of DNA and GFP–DNA strands were assembled and analyzed by gel electrophoresis (Figure 2). In these experiments, conjugates were added so

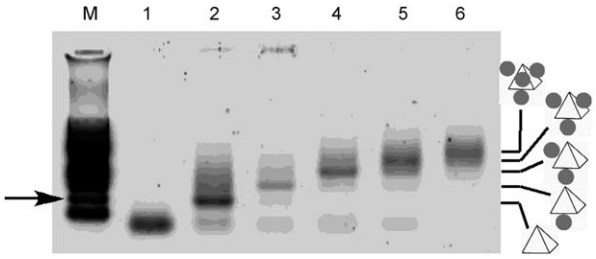


Figure 2. Formation of protein–DNA tetrahedra that contain varying numbers of protein–DNA strands. Lane M: 100-bp DNA ladder, 200 bp is indicated by the arrow; lane 1: oligonucleotide **6**; lane 2: DNA tetrahedron alone (**6**, **7**, **8**, and **9**); lane 3: **10c**, **7**, **8**, and **9**; lanes 4–6: increasing number of GFP–DNA strands. The band assignments are schematically represented on the right-hand side of the figure.

that the final concentration of each species was 100 nM in 200 μL of TM buffer solution. Lane 1 contains **6** and lane 2 contains oligonucleotides **6**, **7**, **8**, and **9** (with no GFP present). Lane 3 contains one GFP–oligonucleotide conjugate (**10c**) and **7**, **8**, and **9**. Subsequent lanes contain additional GFP–oligonucleotides substituted for the oligonucleotide alone. A clear trend can be seen in which an increasing number of protein molecules proportionally decreases the mobility of the complex, indicating that all four GFP–oligonucleotides are required to form the superstructure.

Finally, the GFP content of the putative GFP–DNA tetrahedron was evaluated by using fluorescence correlation spectroscopy, a single-molecule technique.[37] By measuring the fluorescence intensity fluctuations of fluorescent particles passing through a small two-photon observation volume at low concentrations, the number of fluorophores per DNA tetrahedron was quantified through brightness analysis.[38] Brightness is a measure of the fluorescence signal of a single molecule.[39] Consequently, if a particle contains two GFP molecules, the brightness of the complex is twice that of GFP alone.[40] Measurement of the brightness of the DNA tetrahedron (see the Supporting Information) showed that its value increased by equal amounts as the number of GFP-containing strands was increased from one (complex of **10c**, **7**, **8**, and **9**) to four (complex of **10c**, **11c**, **12c**, and **13c**). These experiments provide convincing evidence for the presence of a single molecular entity with four GFP moieties appended to the DNA tetrahedron.

In summary, we have developed a novel chemoenzymatic method to covalently link any protein of interest with DNA and have demonstrated its utility by preparing a complex protein–DNA tetrahedron. As the tetrahedron is composed of four individual DNA strands with unique sequences, this structure contains four individually addressable positions for protein functionalization. The method for protein–DNA conjugation reported herein requires a recognition tag (CVIA) that is quite small when compared with biotin–avidin and other fusion-protein approaches. This allows the additional protein sequence necessary for selective labeling to be minimized and thereby permits more-compact protein spacing and greater control over the integration of proteins within DNA scaffolds. Ultimately, this should facilitate the preparation of a myriad of novel protein–DNA hybrid structures with useful structural and functional features.

Experimental Section

GFP–CVIA was purified and enzymatically labeled with **1** to produce GFP–N_3 as described previously.[34] All DNA synthesis reagents were purchased from Glen Research and all DNA strands were synthesized by using a PerSeptive Biosystems Expedite DNA Synthesizer. DNA modified with Texas Red was purchased from Integrated DNA Technologies, Inc. After the DNA synthesis was complete, the alkyne phosphoramidite was manually coupled to the 5′ end of the DNA that was still attached to the CPG resin. The modified DNA was deprotected, cleaved from the resin, and purified by reversed-phase HPLC.

To synthesize the GFP–DNA conjugates, GFP–N_3 (20 μM) was incubated with alkyne-modified DNA (22 μM), $CuSO_4$ (1.0 mM), TCEP (1.0 mM) and TBTA ligand (100 μM) in a total volume of 500 μL of 50 mM NaH_2PO_4, pH 7.3 buffer solution (phosphate buffer solution; PB). The reaction was maintained at room temperature for 1 h, after which it was passed through a NAP-5 column (Amersham) equilibrated in PB buffer solution to remove $CuSO_4$, TCEP, and TBTA. The GFP–DNA conjugates were purified by using a Superdex-75 gel filtration column (Amersham) in PB buffer solution containing 1.0 M NaCl (phosphate-buffered saline; PBS).

For GFP–TR studies, 100 nM GFP–DNA (in 500 μL of PBS buffer solution) was mixed with varying equivalents of TR–DNA. Fluorescence spectra were acquired by using a Varian Eclipse Fluorescence detector with a λ_{ex} of 488 nm.

For the tetrahedra experiments shown in Figure 1, DNA strands **6**, **7**, **8**, and **9** and GFP–DNA strands **10c**, **11c**, **12c**, and **13c** were prepared with 50 mM Tris-HCl, pH 8.0 and 5.0 mM $MgCl_2$ (TM buffer solution) so that the final concentration of each component was 1.0 μM. The 20 μL reactions were heated to 54 °C for 3 min followed by cooling to 4 °C over approximately 2 min. The samples were diluted with 5× native gel loading buffer solution and separated by using a 1.5% agarose gel. The gels were subject to electrophoresis in Tris, acetate, ethylenediamine tetraacetic acid (TAE) buffer solution at 100 mA for 40 min at 4 °C. For the samples shown in Figure 2, the conjugates were added so that the final concentration of each species was 100 nM in 200 μL of TM buffer solution. After the tetrahedra were formed as described above, the 200-μL samples were concentrated to approximately 30 μL at 4 °C. These samples were diluted with 5× native gel loading buffer solution and separated by using a 1.5% agarose gel as described above. For the samples shown in Figure 1, the gel was first imaged by using a Molecular Dynamics Storm fluorescence flatbed densitometer and then soaked in EtBr (0.5 μg mL^{-1}) in water for 30 min and imaged by using a Bio-Rad Molecular Imager FX phosphorimager. For Figure 2, EtBr (0.5 μg mL^{-1}) was included in the gel itself.

Received: May 3, 2007
Revised: August 8, 2007
Published online: October 12, 2007

Keywords: click reactions · enzymes · nanostructures · protein labeling · protein structures

[1] J. H. Chen, N. C. Seeman, *Nature* **1991**, *350*, 631.
[2] R. P. Goodman, I. A. T. Schaap, C. F. Tardin, C. M. Erben, R. M. Berry, C. F. Schmidt, A. J. Turberfield, *Science* **2005**, *310*, 1661.
[3] Y. Zhang, N. C. Seeman, *J. Am. Chem. Soc.* **1994**, *116*, 1661.
[4] W. M. Shih, J. D. Quispe, G. F. Joyce, *Nature* **2004**, *427*, 618.
[5] E. Winfree, F. Liu, L. A. Wenzler, N. C. Seeman, *Nature* **1998**, *394*, 539.
[6] 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.
[7] H. Yan, S. H. Park, G. Ginkelstein, J. H. Reif, T. H. LaBean, *Science* **2003**, *301*, 1882.
[8] Y. He, Y. Chen, H. P. Liu, A. E. Ribbe, C. D. Mao, *J. Am. Chem. Soc.* **2005**, *127*, 12202.
[9] E. R. Hendrickson, T. M. H. Truby, R. D. Joerger, E. R. Majarian, R. C. Ebersole, *Nucleic Acids Res.* **1995**, *23*, 522.
[10] T. Sano, C. L. Smith, C. R. Cantor, *Science* **1992**, *258*, 120.
[11] I. Burbulis, K. Yamaguchi, A. Gordon, R. Carlson, R. Brent, *Nat. Methods* **2005**, *2*, 31.
[12] S. H. Park, P. Yin, Y. Liu, J. H. Reif, T. H. LaBean, H. Yan, *Nano Lett.* **2005**, *5*, 729.
[13] C. M. Niemeyer, J. Koehler, C. Wuerdemann, *ChemBioChem* **2002**, *3*, 242.
[14] P. Hazarika, F. Kukolka, C. M. Niemeyer, *Angew. Chem.* **2006**, *118*, 6981; *Angew. Chem. Int. Ed.* **2006**, *45*, 6827.
[15] C. M. Niemeyer, T. Sano, C. L. Smith, C. R. Cantor, *Nucleic Acids Res.* **1994**, *22*, 5530.
[16] C. M. Niemeyer, W. Bürger, J. Peplies, *Angew. Chem.* **1998**, *110*, 2391; *Angew. Chem. Int. Ed.* **1998**, *37*, 2265.
[17] C. M. Niemeyer, M. Adler, S. Gao, L. Chi, *Bioconjugate Chem.* **2001**, *12*, 364.
[18] C. M. Niemeyer, C. A. Mirkin, *Nanobiotechnology: Concepts, Methods and Applications*, Wiley-VCH, Weinheim, **2004**.
[19] Y. He, Y. Tian, A. E. Ribbe, C. Mao, *J. Am. Chem. Soc.* **2006**, *128*, 12664.
[20] J. M. Tomkins, B. K. Nabbs, K. Barnes, M. Legido, A. J. Blacker, R. A. McKendry, C. Abell, *ChemBioChem* **2001**, *2*, 375.
[21] H. Li, S. H. Park, J. H. Reif, T. H. LaBean, H. Yan, *J. Am. Chem. Soc.* **2004**, *126*, 418.
[22] Y. Liu, C. Lin, H. Li, H. Yan, *Angew. Chem.* **2005**, *117*, 4407; *Angew. Chem. Int. Ed.* **2005**, *44*, 4333.
[23] S. S. Ghosh, P. M. Kao, A. W. McCue, H. L. Chappelle, *Bioconjugate Chem.* **1990**, *1*, 71.
[24] D. R. Corey, P. G. Schultz, *Science* **1987**, *238*, 1401.
[25] S. Howorka, S. Cheley, H. Bayley, *Nat. Biotechnol.* **2001**, *19*, 636.
[26] F. Kukolka, C. M. Niemeyer, *Org. Biomol. Chem.* **2004**, *2*, 2203.
[27] M. Lovrinovic, R. Seidel, R. Wacker, H. Schroeder, O. Seitz, M. Engelhard, R. S. Goody, C. M. Niemeyer, *Chem. Commun.* **2003**, 822.
[28] M. Lovrinovic, C. M. Niemeyer, *ChemBioChem* **2007**, *8*, 61.
[29] F. Kukolka, B. M. Müller, S. Paternoster, A. Arndt, C. M. Niemeyer, C. Bräuchle, D. C. Lamb, *Small* **2006**, *2*, 1083.
[30] F. Kukolka, O. Schoeps, U. Woggon, C. M. Niemeyer, *Bioconjugate Chem.* **2007**, *18*, 621.
[31] Y. Kho, S. C. Kim, C. Jiang, D. Barma, S. W. Kwon, J. Cheng, J. Jaunbergs, C. Weinbaum, F. Tamanoi, J. Falk, Y. Zhao, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12479.
[32] M. W. Rose, N. D. Rose, J. Boggs, S. Lenevich, J. Xu, G. Barany, M. D. Distefano, *J. Pept. Res.* **2005**, *65*, 529.
[33] C. Gauchet, G. R. Labadie, C. D. Poulter, *J. Am. Chem. Soc.* **2006**, *128*, 9274.
[34] B. P. Duckworth, J. Xu, T. A. Taton, A. Guo, M. D. Distefano, *Bioconjugate Chem.* **2006**, *17*, 967.
[35] R. P. Goodman, R. M. Berry, A. J. Turberfield, *Chem. Commun.* **2004**, 1372.
[36] C. M. Erben, R. P. Goodman, A. J. Turberfield, *Angew. Chem.* **2006**, *118*, 7574; *Angew. Chem. Int. Ed.* **2006**, *45*, 7414.
[37] E. L. Elson, *J. Biomed. Opt.* **2004**, *9*, 857.
[38] Y. Chen, J. D. Mueller, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 3147.
[39] Y. Chen, J. D. Mueller, P. T. C. So, E. Gratton, *Biophys. J.* **1999**, *77*, 553.
[40] Y. Chen, L.-N. Wei, J. D. Mueller, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15492.