COMMUNICATIONS

J|A|C|S

Multistep Small-Molecule Synthesis Programmed by DNA Templates

Zev J. Gartner, Matthew W. Kanan, and David R. Liu*

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Chemists generate molecules with new function by synthesizing structures, assaying their properties, formulating structure−activity relationships, and then synthesizing new candidates. In contrast, Nature generates functional proteins and nucleic acids through iterated cycles of translation, selection, amplification, and diversification, collectively referred to as molecular evolution. Although the power of Nature's approach is well appreciated,[1] we cannot apply molecular evolution to molecules other than proteins or nucleic acids because we lack a method of translating an amplifiable information carrier such as DNA into arbitrary structures. One approach to this problem uses DNA-templated synthesis to translate libraries of DNA sequences into corresponding synthetic molecules. Recently we found[2] that DNA-templated synthesis can direct a variety of powerful chemical reactions with high sequence specificity and without requiring structural mimicry of the DNA backbone. The application of this approach to synthetic molecules of potentially useful complexity, however, requires the development of methods that enable the product of a DNA-templated reaction to undergo subsequent DNA-templated transformations. Here we report the first DNA-templated multistep small-molecule syntheses. Our results represent a general approach to translating DNA sequences into corresponding multistep synthetic products in a form suitable for their in vitro evolution.

The requirements of DNA-templated synthesis for in vitro selection and amplification differ from those of traditional or combinatorial synthesis. Overall yields are much less important than in current approaches to synthesis because as little as $\sim 10^{-20}$ mol of template is needed for PCR amplification[3] following selection. The sequence specificity of DNA-templated reactions, however, is crucial to maintaining a 1:1 correspondence between the information encoded in the template and the structure of the covalently linked synthetic molecule. Multistep DNA-templated small-molecule synthesis faces two additional challenges. First, the DNA used to direct reagents to appropriate templates must be removed from the product of a DNA-templated reaction prior to subsequent DNA-templated steps. Second, multistep synthesis often requires the purification and isolation of intermediate products, yet common methods used to purify reaction products are not suitable for synthesis on the molecular biology (picograms to micrograms) scale.

To address the first of these challenges, we implemented three distinct strategies for linking chemical reagents with their decoding DNA oligonucleotides, inspired by previous work[4] in solid-phase organic synthesis. In the "scarless linker" approach, one functional group of a reagent is reserved for DNA-templated bond formation, while a second functional group is used to attach a linker that can be cleaved without introducing additional unwanted chemical functionality. For example, we synthesized aminoacylation reagents such as L-Phe derivative **1**, in which the α-amine is connected to its decoding DNA oligonucleotide through a carbamoylethyl sulfone

linker.[4c] The product (**2**) of DNA-templated amide bond formation[2b] using this reagent and an amine-terminated template (**3**) was treated with pH 11.8 buffer to effect the quantitative elimination and spontaneous decarboxylation of the linker, affording product **4** containing the cleanly transferred amino acid (Figure 1).
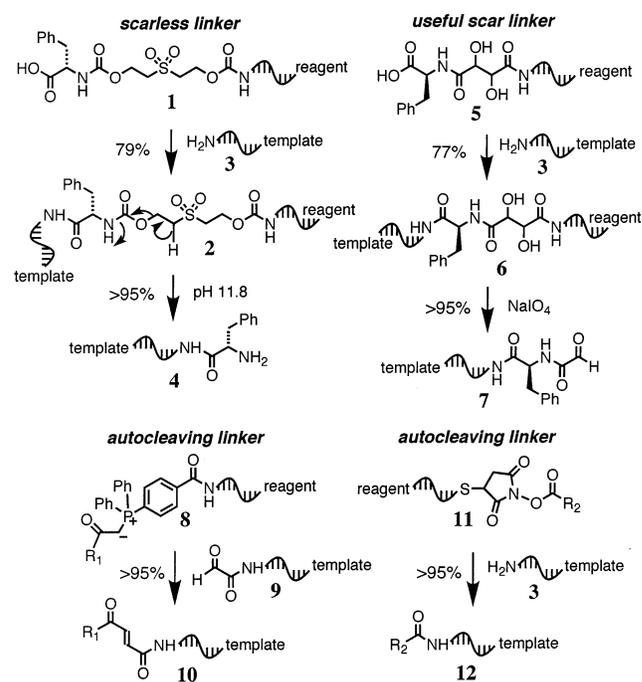

Figure 1. Three linker strategies for DNA-templated synthesis. $R_1 =$ $NH(CH_2)_2NH$-dansyl; $R_2 =$ biotinyl. Reaction yields were quantitated by denaturing polyacrylamide gel electrophoresis followed by densitometry (see Supporting Information).

In some cases it may be advantageous to introduce new chemical groups as a consequence of linker cleavage. Under a second linker strategy, cleavage generates a "useful scar" that can be functionalized in subsequent steps. As an example of this class of linker, we generated amino acid reagents such as L-Phe derivative **5** linked through 1,2-diols[4d] to their decoding DNA oligonucleotides. Following DNA-templated amide bond formation with amine-terminated template **3** to provide amide **6**, this linker was quantitatively cleaved by oxidation with 50 mM aqueous $NaIO_4$ to afford product **7** containing an aldehyde group appropriate for subsequent functionalization (for example, in a DNA-templated Wittig olefination, reductive amination, or nitro-aldol addition[2b]) (Figure 1).

When possible, an ideal reagent−oligonucleotide linker for DNA-templated synthesis positions the oligonucleotide as a leaving group of the reaction. Under this third linker strategy, the oligonucleotide− reagent bond is cleaved as a natural consequence of the reaction.

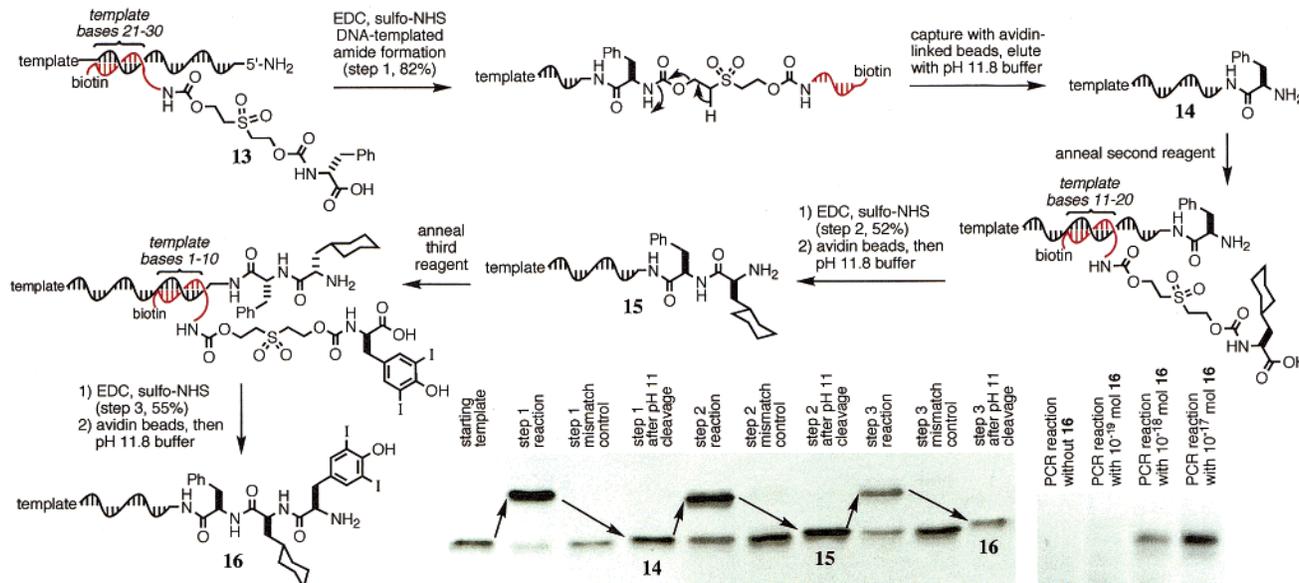* Address correspondence to this author. E-mail: drliu@fas.harvard.edu.

Figure 2. DNA-templated multistep tripeptide synthesis. The progress of each reaction and purification was followed by denaturing polyacrylamide gel electrophoresis (bottom) and, in the case of final product **16**, by MALDI mass spectrometry (see text). Lanes 3, 6, and 9: control reactions using reagents containing scrambled oligonucleotide sequences. Lanes 11−14: PCR reactions using the listed amounts of final product **16** as a template. EDC, 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide.
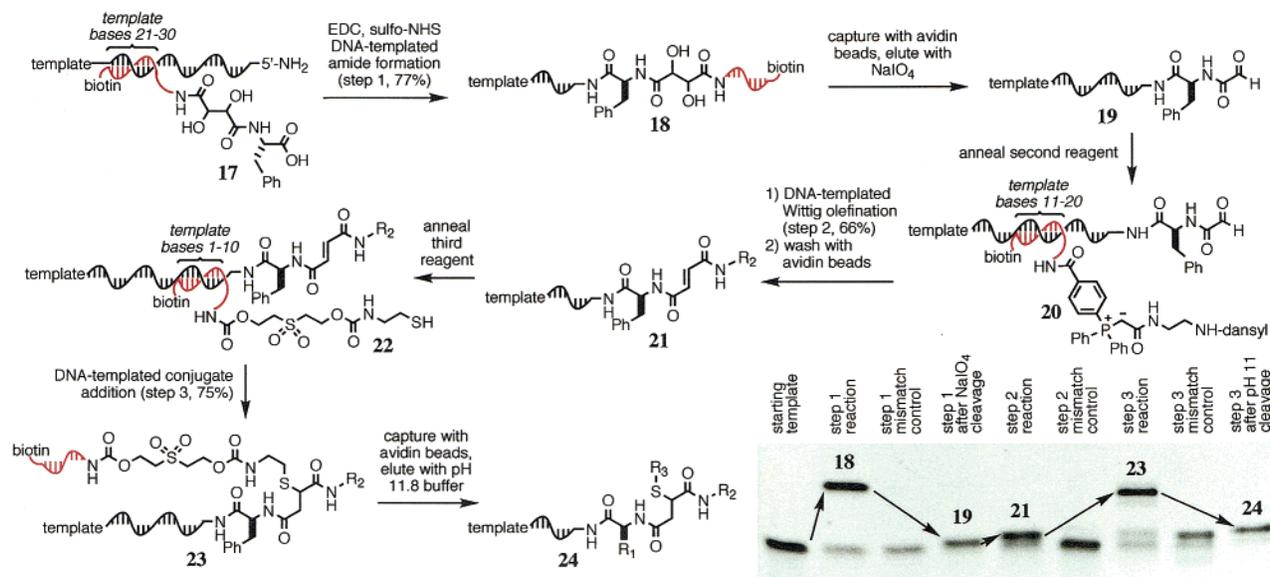

Figure 3. DNA-templated multistep synthesis of a branched thioether. The progress of each reaction and purification was followed by denaturing polyacrylamide gel electrophoresis (bottom) and, in the case of **24**, by MALDI mass spectrometry (see text). Lanes 3, 6, and 9: control reactions using reagents containing scrambled oligonucleotide sequences. The olefin geometry of **21** is presumed, and only one of the possible thiol addition regioisomers is shown in **23** and **24**. $R_1 = CH_2Ph$; $R_2 = (CH_2)_2NH$-dansyl; $R_3 = (CH_2)_2NH_2$.

As one example of this "autocleaving linker" strategy, we synthesized a Wittig phosphorane reagent (**8**) in which the decoding DNA oligonucleotide was attached through a triphenyl phosphine group.[4b] DNA-templated Wittig olefination[2b] with aldehyde-linked template **9** resulted in nearly quantitative, sequence-specific olefination to provide **10** unattached to reagent oligonucleotide (Figure 1). As a second example of an autocleaving linker, DNA-linked *N*-hydroxysuccinimide (NHS) ester **11** sequence-specifically acylated amino-terminated template **3** to quantitatively afford amide **12** (Figure 1) in an approach similar to that used by Nature during ribosomal protein biosynthesis.

To address the second challenge of product purification, we have developed two general approaches for purification after any DNA-templated synthetic step using biotinylated reagent oligonucleotides.

For reactions using the scarless or useful scar linker strategies, unreacted reagents together with desired products (covalently linked to biotinylated oligonucleotides) are captured on streptavidin-linked magnetic beads. Any unreacted template bound to reagent by base pairing is removed by washing the beads with buffer containing 4 M guanidinium chloride. Biotinylated molecules remain bound to the streptavidin beads under these conditions. Desired product is then isolated in pure form by eluting the beads with linker cleavage buffer, while reacted and unreacted reagents remain bound to the beads (Figures 2 and 3). For autocleaving DNA-templated reactions, desired products can be purified away from unreacted reagents and cleaved oligonucleotides by washing crude reactions with strepta-vidin-linked magnetic beads. Although this purification approach does not separate reacted templates from unreacted templates, the

latter can be removed by gel purification or in subsequent reaction and purification steps (Figure 3).

Integrating the recently expanded repertoire of DNA-templated reactions[2] with the linker and purification strategies described above, we conducted two multistep DNA-templated small-molecule syntheses. In the first example, three cycles of DNA-templated amide formation,[2b] scarless linker cleavage, and purification with streptavidin-linked beads were used to generate a non-natural tripeptide (Figure 2). Each amino acid reagent was linked to a unique biotinylated 10-base DNA oligonucleotide through the sulfone linker described above. The 30-base amine-terminated template programmed to direct the tripeptide synthesis contained three consecutive 10-base regions that were complementary to the reagent oligonucleotides, mimicking the strategy that would be used in a multistep DNA-templated library synthesis. The first amino acid reagent (13) was activated and combined with the amino-terminated template to effect DNA-templated amide formation. The desired product was purified by capturing with streptavidin-linked magnetic beads and eluting with pH 11.8 buffer to effect sulfone linker cleavage, providing product 14 with a new free amine group. A second and third round of DNA-templated amide formation and linker cleavage afforded dipeptide 15 and tripeptide 16 (Figure 2).

Each reaction, purification, and linker cleavage step was followed by denaturing polyacrylamide gel electrophoresis. The yields of each DNA-templated amide formation ranged from 52% to 82% (Figure 2). Product was not formed when the starting template was capped with acetic anhydride, or when reagents containing sequence mismatches were used instead of the matched reagents (Figure 2). The desired tripeptide linked to its encoding template was isolated in 3% overall yield for three DNA-templated steps, three purification steps, and three linker cleavage steps. The final tripeptide linked to template (16) was characterized by MALDI mass spectrometry (expected mass, 10 069; observed mass, 10 067 ± 8). As little as 10^{-18} mol of 16 was successfully amplified by PCR (Figure 2). Beginning with a modest quantity of starting material (2.5 nmol, 25 µg), sufficient final product 16 was therefore generated to serve as the template for 10^7 PCR reactions, or, in theory, 1000 different in vitro selections followed by PCR amplification if only one in 10 000 molecules survives selection.

We also performed a multistep DNA-templated small-molecule synthesis that uses all three linker strategies developed above together with three different DNA-templated reactions (Figure 3). An amine-terminated 30-base template was subjected to DNA-templated amide formation[2b] using an aminoacyl donor reagent (17) containing the diol linker and a biotinylated 10-base oligonucleotide to afford amide 18. The desired product was purified by capturing the crude reaction on streptavidin beads, followed by cleaving the linker with NaIO$_4$ to provide aldehyde 19. The DNA-templated Wittig reaction[2b] of 19 with the biotinylated autocleaving phosphorane reagent 20 afforded fumaramide 21. The products from the second DNA-templated reaction were partially purified by washing with streptavidin beads to remove reacted and unreacted reagent. In the third step, fumaramide 21 was subjected to a DNA-templated conjugate addition[2a] using sulfone-linked thiol reagent 22. The conjugate addition product (23) was purified by immobilization with streptavidin beads and then cleaved with pH 11.8 buffer to provide final product 24 (a mixture of possible regioisomers and enantiomers). Yields of each DNA-templated reaction ranged from 66% to 77% (Figure 3). Branched thioether 24 was made in 5−10% overall isolated yield for three bond-forming reactions, two linker cleavage steps, and three purifications.

This final product was digested with EcoRI, and the mass of the small-molecule-linked template fragment was confirmed by MALDI mass spectrometry (expected mass, 2568; observed mass, 2564.3 ± 5). As in the tripeptide example, control reactions with sequence mismatches yielded no product (Figure 3). Control reactions in which the Wittig reagent was omitted in step 2 also did not generate any products during the third step, consistent with olefin formation in step 2. Taken together, the DNA-templated syntheses of 16 and 24 demonstrate the ability of DNA to direct the sequence-programmed multistep synthesis of small molecules unrelated in structure to nucleic acids.

Together with the diverse collection of sequence-specific DNA-templated synthetic reactions described in previous reports,[2,5] the development of multistep synthesis programmed by DNA templates brings the evolution of synthetic molecules closer to realization. In addition, multistep nucleic acid-templated synthesis is a requirement of proposed models[6] for the prebiotic translation of replicable information into functional molecules. Our findings demonstrate that nucleic acid templates are indeed capable of directing multistep small-molecule synthesis even when reagents anneal at widely varying distances from the growing molecule (in the above examples, 0−20 bases). The use of methods developed here to evolve libraries of synthetic molecules toward active ligands and catalysts through cycles of translation, selection, amplification, and mutagenesis is underway in our laboratory.

Supporting Information Available: Experimental details and gel electrophoresis images of the linker cleavage reactions shown in Figure 1 (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Arnold, F. H.; Volkov, A. A. *Curr. Opin. Chem. Biol.* **1999**, *3*, 54−59. (b) Jaschke, A.; Seelig, B. *Curr. Opin. Chem. Biol.* **2000**, *4*, 257−262.

(2) (a) Gartner, Z. J.; Liu, D. R. *J. Am. Chem. Soc.* **2001**, *123*, 6961−6963. (b) Gartner, Z. J.; Kanan, M. W.; Liu, D. R. *Angew. Chem., Int. Ed.* **2002**, *41*, 1796−1800.

(3) Kramer, M. F.; Coen, D. M. In *Current Protocols in Molecular Biology*; Ausubel, F. M., Ed.; Wiley: 1999; Vol. 3, p 15.11−15.13.

(4) (a) Guillier, F.; Orain, D.; Bradley, M. *Chem. Rev.* **2000**, *100*, 2091−2157. (b) Hughes, I. *Tetrahedron Lett.* **1996**, *37*, 7595−7598. (c) Zarling, D. A.; Watson, A.; Bach, F. H. *J. Immunol.* **1980**, *124*, 913−920. (d) Fruchart, J.-S.; Gras-Masse, H.; Melnyk, O. *Tetrahedron Lett.* **1999**, *40*, 6225−6228.

(5) (a) Schmidt, J. G.; Christensen, L.; Nielsen, P. E.; Orgel, L. E. *Nucleic Acids Res.* **1997**, *25*, 4792. (b) Bruick, R. K.; Dawson, P. E.; Kent, S. B. H.; Usman, N.; Joyce, G. F. *Chem. Biol.* **1996**, *3*, 49−56. (c) Czlapinski, J. L.; Sheppard, T. L. *J. Am. Chem. Soc.* **2001**, *123*, 8618−8619. (d) Li, X.; Zhan, Z. Y.; Knipe, R.; Lynn, D. G. *J. Am. Chem. Soc.* **2002**, *124*, 746−747. (e) Xu, Y.; Kool, E. T. *Nucleic Acids Res.* **1999**, *27*, 875−881. (f) Herrlein, M. K.; Nelson, J. S.; Letsinger, R. L. *J. Am. Chem. Soc.* **1995**, *117*, 10151−10152. (g) Orgel, L. E. *Acc. Chem. Res.* **1995**, *28*, 109−118. (h) Luther, A.; Brandsch, R.; von Kiedrowski, G. *Nature* **1998**, *396*, 245−248.

(6) (a) Scott, A. I. *Tetrahedron Lett.* **1997**, *38*, 4961−4964. (b) Li, T.; Nicolaou, K. C. *Nature* **1994**, *369*, 218−220. (c) Tamura, K.; Schimmel, P. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 1393−1397.

JA027307D