J|A|C|S
COMMUNICATIONS

Published on Web 00/00/0000

Efficient and Sequence-Specific DNA-Templated Polymerization of Peptide Nucleic Acid Aldehydes

Daniel M. Rosenbaum and David R. Liu*

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

Received August 21, 2003; E-mail: drliu@fas.harvard.edu

The highly efficient and sequence-specific nucleic acid-templated polymerization of proteins and nucleic acids is a fundamental feature of living systems that enables biopolymers to evolve.[1] Over the past few decades, several efforts to adapt templated polymerization to synthetic systems have been reported.[2] These systems typically use oligonucleotides (or analogues) as a template for coupling DNA- or RNA-based monomers. Despite significant advances,[3] the generality and sequence specificity of template-directed polymerization is still largely unexplored. For example, the efficient and sequence-specific templated polymerization of easily functionalized synthetic monomers lacking a ribose backbone has not been reported. Such a system may enable polymers comprised of these synthetic monomers to evolve through iterated cycles of translation (polymerization), selection, and amplification presently available only to DNA, RNA, and proteins.

The minimal requirements of a system for synthetic polymer evolution are (i) efficient nucleic acid-templated oligomerization to provide sufficient quantities of full-length products for in vitro selections;[4] (ii) distance-dependent nucleic acid-templated monomer coupling reactions[5] to ensure that oligomerization proceeds exclusively between adjacently annealed monomers; (iii) stable linkage of each synthetic polymer to its encoding nucleic acid template to ensure the survival of the appropriate template during polymer selection; and (iv) a readily functionalized synthetic monomer backbone to introduce chemical functionality into the polymer.

Peptide nucleic acids (PNAs) are attractive candidates for synthetic polymer evolution because of their known ability to bind DNA sequence specifically and their simple preparation from synthetically accessible amino acids. Previous efforts to oligomerize PNAs on DNA or RNA templates used amine acylation as the coupling reaction and proceeded with modest efficiency and sequence specificity.[6] Consistent with these findings, when we combined five PNA tetramers using a variety of aqueous amine acylation conditions[5] in the presence of DNA templates containing complementary 20-base annealing regions, we observed only modest formation ($\leq$20% yield) of full-length PNAs representing five successive coupling reactions (Supporting Information). Even more problematic, the formation of higher molecular weight products was observed independent of the position of a mismatched 4-base annealing region in the template (Supporting Information). These observations indicate that PNAs are able to couple using amine acylation chemistry even when not adjacently annealed, leading to an unpredictable mixture of products.

We hypothesized that the distance independence we previously reported in DNA-templated amine acylation reactions[5b,c] was the origin of the poor regiospecificity of amine acylation-mediated PNA couplings. A distance-dependent DNA-templated reaction such as reductive amination[5b,c] should address this problem and serve as a more promising chemistry for DNA-templated PNA polymerizations. Indeed, Lynn and co-workers previously reported excellent

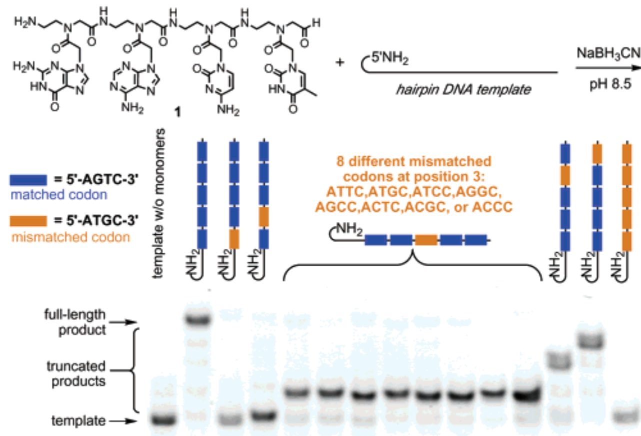

Figure 1. DNA-templated PNA aldehyde oligomerization. A 5′-amine-terminated DNA template (0.4 μM) was mixed with PNA aldehyde **1** (8 μM), heated to 95 °C, and cooled to 25 °C over 60 min. $NaCNBH_3$ was added to 80 mM. After 30 min, reactions were subjected to gel filtration and denaturing PAGE (bottom). Band doubling in lanes 13 and 14 is attributed to open versus closed hairpin structures in the products.

results with reductive amination-mediated step-growth oligomerization of DNA analogues.[3b]

A thymine-containing PNA monomer amino aldehyde was synthesized and coupled to threonine-linked resin.[7] Standard FMOC peptide synthesis was used to extend the peptide by three PNA monomers (final sequence: NH_2−gact−CHO), and aqueous acidic cleavage from the resin yielded the desired tetrameric peptide aldehyde (**1**, Figure 1). A DNA template containing a 5′-amine-terminated hairpin and five successive repeats of the 5′-AGTC-3′ "codon" complementary to **1** was combined with 8 μM **1** in aqueous pH 8.5 buffer. The reactants were annealed, and $NaCNBH_3$ was added. The starting template was almost entirely consumed, and a higher molecular weight product was formed in >90% yield (Figure 1). Gel purification of the product followed by removal of the DNA template with DNase I and MALDI-TOF mass spectrometry confirmed a full-length pentamer of the gact PNA aldehyde (expected mass = 5366; observed mass = 5375 $\pm$ 15). No product was observed in the absence of $NaCNBH_3$ (Supporting Information). These results indicate that DNA-templated reductive amination can mediate the highly efficient oligomerization of PNA aldehydes.

To examine the regio- and sequence specificity of this reaction, we repeated the above oligomerization using a variety of template sequences. Remarkably, when a mismatched DNA template codon (5′-ATGC-3′) was introduced at the second, third, fourth, or fifth 4-base coding region of the template, highly efficient (>90% yield) formation of products corresponding to the coupling of one, two, three, or four copies of **1**, respectively, was observed (Figure 1). When the mismatched codon was placed at only the first coding position, or at all five coding positions, no product formation was observed (Figure 1). The termination of oligomerization at the first mismatched codon in every case indicates that this DNA-templated

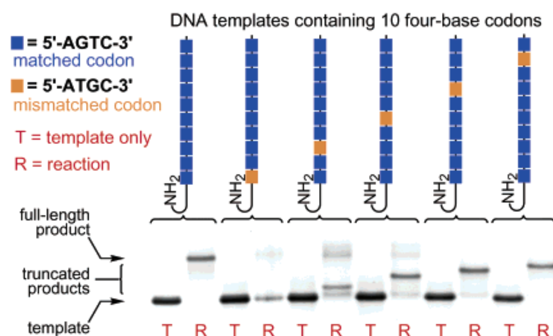

Figure 2. Polymerization on long (40-base) templates. Polymerizations were carried out as in Figure 1, with the following exceptions: PNA aldehyde concentration was 16 μM, and reaction time with $NaCNBH_3$ was 15 min. Gel lanes alternate between template only (with mismatch at indicated position) and reactions (template + gact monomer).

PNA aldehyde coupling indeed requires functional group adjacency (i.e., is highly distance dependent[5]) and therefore is ideally suited for templated polymerizations.[3b]

We further probed the sequence specificity of this system by performing oligomerization experiments using DNA templates containing eight different mismatched codons (ATTC, ATGC, ATCC, AGGC, AGCC, ACTC, ACGC, or ACCC) in the third coding region. Even though four of these codons differ from the matched sequence (AGTC) in only a single base, in each case only two copies of **1** were coupled to the template (Figure 1). This high degree of sequence specificity raises the possibility that libraries of different DNA sequences may be faithfully translated into libraries of corresponding polymers using this system, analogous to our ongoing DNA-templated small-molecule studies.[5]

Synthetic polymers with desired properties may require lengths beyond those previously achieved efficiently using nucleic acid-templated synthesis. To test the ability of the above system to generate longer polymers efficiently and sequence specifically, we translated DNA templates with 40-base coding regions encoding 10 repeats of the above matched or mismatched codon into corresponding PNA aldehyde polymers. Gratifyingly, both denaturing PAGE and MALDI-TOF mass spectrometry revealed a single predominant product corresponding to the polymerization of a full length 40-mer PNA after 15 min (expected mass = 10 719; observed mass = 10 729 ± 30) (Figure 2 and Supporting Information). Introducing a mismatched codon in the first, third, fifth, seventh, or ninth coding positions on the template again resulted in truncation (Figure 2). This efficient translation of DNA sequences into 40 PNA bases (10 couplings) provides a polymer of length similar to that of DNA and RNA oligonucleotides with binding or catalytic properties,[4a] but made entirely of synthetic building blocks.

A challenging requirement of creating libraries of sequence-defined synthetic polymers in this manner is maintaining sequence specificity in the presence of multiple monomers of closely related sequence. To study the specificity of DNA-templated polymerization using multiple PNA building blocks in a single solution, we synthesized the nine PNA aldehyde tetramers of the sequence NH_2−gvvt−CHO (v = g, a, or c). We also prepared nine DNA templates containing one of nine codons complementary to gvvt at codon 5, and containing AGTC at codons 1−4 and 6−10. Each of the eight templates not containing AGTC at position 5 was translated into a predominant truncated product of one apparent length when **1** was the only PNA aldehyde included in the reaction (Figure 3, 0 lanes). Full-length polymer was the major product for all nine templates, however, when the PNA aldehyde complementary to the fifth codon was included in addition to **1** (C lanes), or when all nine PNA aldehydes were included (A lane). Importantly, when all PNA

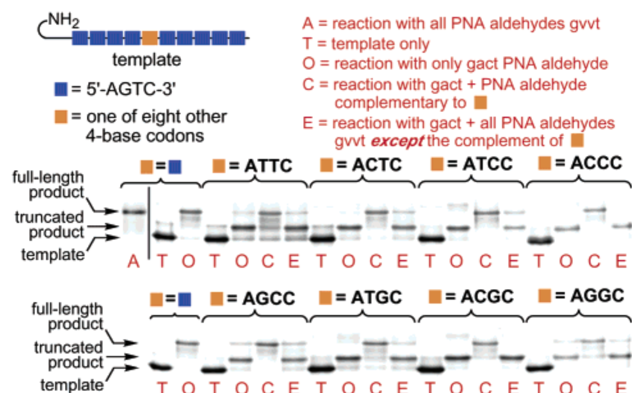

Figure 3. Polymerization in the presence of multiple PNA aldehydes. Conditions are the same as those in Figure 2, except that the reaction time was 5 min at 37 °C. Lane 1 (A) contains 16 μM **1** + 2 μM of each of the eight other gvvt PNA aldehydes. Each set of four lanes after lane 3 contains: (T) template only; (O) reaction with 16 μM **1**; (C) reaction with 14.4 μM **1** + 1.6 μM PNA aldehyde complementary to the orange codon; (E) reaction with 14.4 μM **1** + 0.2 μM of each PNA aldehyde of sequence gvvt except the PNA complementary to the orange codon (eight PNAs total).

aldehyde tetramers were included in the reaction *except* the PNA complementary to the fifth coding region (eight in total), again only the truncated product was predominantly generated (Figure 3, E lanes). Taken together, these experiments reveal that DNA-templated PNA aldehyde polymerizations maintain efficiency and sequence specificity even in the presence of a mixture of different PNA building blocks closely related in sequence.

In summary, we have developed an efficient and sequence-specific translation of DNA templates containing as many as 40 bases of mixed coding sequence into corresponding synthetic peptide nucleic acid polymers. These findings are a key step toward the evolution of sequence-defined synthetic heteropolymers, rather than biological macromolecules, using iterated cycles of DNA-templated translation, in vitro selection, PCR amplification, and sequence diversification.

Acknowledgment. This work was supported by the Office of Naval Research (N00014-03-1-0749), the Beckman Foundation, the Searle Scholars Program (00-C-101), and the Sloan Foundation (BR-4141). D.M.R. is a Hertz Foundation Graduate Fellow. We thank Chieh-Ting Yeh for experimental assistance.

Supporting Information Available: Experimental details, oligonucleotide sequences, and additional results (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Kuchta, R. D.; Benkovic, P.; Benkovic, S. J. *Biochemistry* **1988**, *27*, 6716−6725. (b) Steitz, T. A. *Harvey Lect.* **1997−98**, *93*, 75. (c) Yonath, A. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 257.
(2) (a) Joyce, G. F. *Nonezymatic Template-Directed Synthesis of Informational Macromolecules*; Cold Spring Harbor Laboratory Press: Plainview, NY, 1987; LII, pp 41−51. (b) Orgel, L. E. *Nature* **1992**, *358*, 203−209. (c) Bag, B. G.; von Keidrowski, G. *Pure Appl. Chem.* **1996**, *68*, 2145−2152. (d) Gat, Y.; Lynn, D. G. *Templated Organic Synthesis*; Wiley: New York, 2000; p 133.
(3) (a) Kozlov, I. A.; De Bouvere, B.; Aerschot, A. V.; Herdewijn, P.; Orgel, L. E. *J. Am. Chem. Soc.* **1999**, *121*, 5856−5859. (b) Li, X.; Zhan, Z. J.; Knipe, R.; Lynn, D. G. *J. Am. Chem. Soc.* **2002**, *124*, 746−747.
(4) (a) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611−647. (b) Doyon, J. B.; Snyder, T. M.; Liu, D. R. *J. Am. Chem. Soc.* **2003**, *125*, 12372−12373.
(5) (a) Gartner, Z. J.; Liu, D. R. *J. Am. Chem. Soc.* **2001**, *123*, 6961. (b) Gartner, Z. J.; Kanan, M. W.; Liu, D. R. *Angew. Chem., Int. Ed.* **2002**, *41*, 1796. (c) Gartner, Z. J.; Grubina, R.; Calderone, C. T.; Liu, D. R. *Angew. Chem., Int. Ed.* **2003**, *42*, 1370−1375.
(6) (a) Bohler, C.; Nielsen, P. E.; Orgel, L. E. *Nature* **1995**, *376*, 578−581. (b) Schmidt, J. G.; Christensen, L.; Nielsen, P. E.; Orgel, L. E. *Nucleic Acids Res.* **1997**, *25*, 4792−4796.
(7) Ede, N. J.; Bray, A. M. *Tetrahedron Lett.* **1997**, *38*, 7119−7122.

JA038058B