

## Templated Synthesis of Peptide Nucleic Acids via Sequence-Selective Base-Filling Reactions

Jennifer M. Heemstra and David R. Liu\*

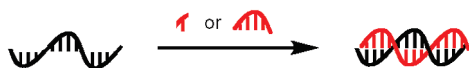
Howard Hughes Medical Institute and the Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138

Received June 9, 2009; E-mail: drliu@fas.harvard.edu

Template-directed nucleic acid synthesis is an essential component of living systems (and their early precursors) that allows genetic information to be passed down to subsequent generations. The desire to understand early nucleic acid replication systems has inspired the search for nonenzymatic routes to templated nucleic acid polymerization in the laboratory.<sup>1</sup> Previous examples of templated nucleic acid synthesis have coupled nucleotide monomers or short nucleotide oligomers by joining backbone functional groups in reactions including, or analogous to, phosphodiester formation (Figure 1a). While these reactions can yield the desired products, they often produce additional nontemplated products and can exhibit modest sequence specificity.<sup>1b</sup>

We envisioned base filling, the addition of individual bases to an abasic backbone (Figure 1b), as an alternative route to the nonenzymatic templated synthesis of nucleic acids.<sup>2,3</sup> Compared with the backbone ligation approach, a base-filling approach offers at least two advantages: (i) the nucleobase monomers do not contain complementary functional groups and therefore are not self-reactive, potentially reducing nontemplated reactions; and (ii) the reaction site is as close as possible to the site of base pairing, potentially increasing sequence specificity. Here we describe the templated synthesis of both native and modified peptide nucleic acids (PNAs) through base filling.

a) templated nucleic acid synthesis by backbone ligation

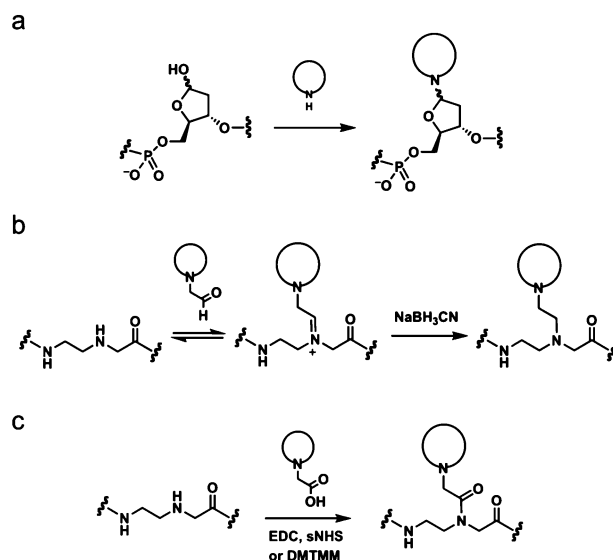


b) templated nucleic acid synthesis by base filling



**Figure 1.** Two strategies for the templated polymerization of nucleic acids: (a) backbone ligation and (b) base filling.

Early hypotheses on the mechanism of prebiotic nucleic acid replication considered the addition of bases to an abasic backbone to be unlikely given the unfavorable equilibrium of *N*-glycosidic bond formation in water,<sup>4,5</sup> although Bartel and co-workers more recently demonstrated the ability of a ribozyme to catalyze the addition of uracil to activated ribose pyrophosphate.<sup>6</sup> Over the past few decades, however, researchers have introduced several nucleic acid analogues, including some for which prebiotic syntheses have been proposed, that may support base filling through reactions more favorable than *N*-glycosidic bond formation. One of the most well studied of these nucleic acid variants is PNA.<sup>7</sup> We considered PNA to be a promising candidate for replication by base filling since the abasic reaction site is a versatile secondary amine. Base filling with nucleobase acetic acids by amine acylation would provide native PNA, and base filling with nucleobase acetaldehydes by reductive amination would generate deoxyPNA (dPNA) (Figure 2).



**Figure 2.** Base-filling chemistries. (a) Base filling of DNA by *N*-glycosidic bond formation. (b) Base filling of PNA by reductive amination to generate dPNA. (c) Base filling of PNA by amine acylation to generate native PNA.

The incorporation of a single dPNA monomer in the middle of a PNA–DNA duplex has been found to dramatically lower  $T_m$  values.<sup>8</sup> When the modification is at the end of the strand, however, there is little effect on  $T_m$ ,<sup>9</sup> likely due to the greater structural flexibility at the end of the duplex. In light of these observations, we designed and synthesized PNA templates and substrates **1–15** that collectively represent the base filling of each of the four nucleobases at a site in the middle of a PNA duplex, as well as at a site at the end of the duplex (Table 1). Dimethyl lysine was added at each terminus to improve water solubility and to prevent aggregation. Template and substrate strands were capped at their *N*-termini with a hexanoyl (Hex) or acetyl (Ac) group, respectively, to make the products and templates distinguishable by mass spectrometry.

With these PNA strands in hand, we conducted a series of  $T_m$  experiments and observed that the dPNA monomer lowered melting temperature significantly when present in the middle of a PNA–PNA duplex ( $\Delta T_m = -9.8$  °C), but to a much lesser extent at the end of the duplex ( $\Delta T_m = -0.9$  °C), consistent with the previous studies on DNA–PNA duplexes.<sup>8,9</sup> We found that incorporation of an abasic site had a parallel, though more dramatic effect on  $T_m$  ( $\Delta T_m = -15.7$  °C when present in the middle of the duplex, and  $\Delta T_m = -3.8$  °C when present at the end of the duplex) (Table 1).

Next we tested the ability of PNA bases to fill a single abasic site in the middle of a PNA strand, where the incoming base benefits from both Watson–Crick base pairing and base-stacking interactions. PNA strands **3**, **4**, **5**, or **6** were combined with their

**Table 1.** Effect of PNA Modifications on Melting Temperature<sup>a</sup>

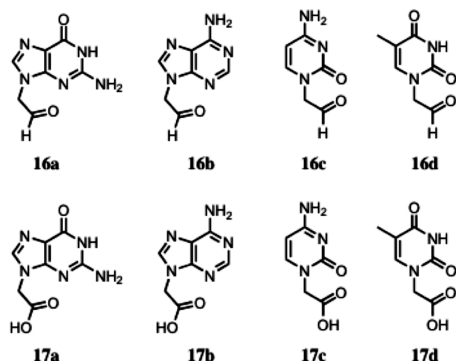
- 1: Hex-Lys(Me<sub>2</sub>)-TAC TGC TAG ACG-Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 2: Hex-Lys(Me<sub>2</sub>)-CGT CTA GCA GTA-Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 3: Ac-Lys(Me<sub>2</sub>)-CGT CTA CA GTA-Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 4: Ac-Lys(Me<sub>2</sub>)-CGT CT GCA GTA-Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 5: Ac-Lys(Me<sub>2</sub>)-TAC TGC AG ACG-Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 6: Ac-Lys(Me<sub>2</sub>)-TAC TG TAG ACG-Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 7: Ac-Lys(Me<sub>2</sub>)-TAC TGC TAG AC Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 8: Ac-Lys(Me<sub>2</sub>)-CGT CTA GCA GT Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 9: Ac-Lys(Me<sub>2</sub>)-AC TGC TAG ACG-Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 10: Ac-Lys(Me<sub>2</sub>)-GT CTA GCA GTA-Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 11: Ac-Lys(Me<sub>2</sub>)-CGT CTA\* GCA GTA-Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 12: Ac-Lys(Me<sub>2</sub>)-CGT CTA GCA GTA\*-Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 13: Ac-Lys(Me<sub>2</sub>)-CGT CT CA GTA-Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 14: Hex-Lys(Me<sub>2</sub>)-TGT ACG CGA TCA-Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 15: Ac-Lys(Me<sub>2</sub>)-TGA TC GT ACA-Lys(Me<sub>2</sub>)-NH<sub>2</sub>  
 — denotes abasic site, \* denotes dPNA monomer

PNA duplex <sup>b</sup>	T <sub>m</sub> (°C)
1 <sup>N</sup> -TAC TGC TAG ACG <sup>C</sup>	81.0
2 <sup>C</sup> -ATG ACG ATC TGC <sup>N</sup>	
1 <sup>N</sup> -TAC TGC TAG ACG <sup>C</sup>	65.3
4 <sup>C</sup> -ATG ACG TC TGC <sup>N</sup>	
1 <sup>N</sup> -TAC TGC TAG ACG <sup>C</sup>	71.2
11 <sup>C</sup> -ATG ACG A*TC TGC <sup>N</sup>	
1 <sup>N</sup> -TAC TGC TAG ACG <sup>C</sup>	77.2
8 <sup>C</sup> -TG ACG ATC TGC <sup>N</sup>	
1 <sup>N</sup> -TAC TGC TAG ACG <sup>C</sup>	80.1
12 <sup>C</sup> -A*TG ACG ATC TGC <sup>N</sup>	

<sup>a</sup> Conditions: 3 mM each PNA, 10 mM potassium phosphate buffer, pH 6.2. <sup>b</sup> The Lys(Me<sub>2</sub>) and capping groups are omitted for clarity.

complementary template (**1** or **2**) and either an equimolar mixture of all four aldehyde monomers **16a–d**<sup>10</sup> under reductive amination conditions, or an equimolar mixture of all four carboxylic acid monomers **17a–d** under amide bond-forming conditions (Figure 3), and the reactions were monitored by MALDI mass spectrometry. Representative MALDI spectra for the selective addition of guanine to **3** by either reductive amination or acylation are shown in Figure 4. All reactions were performed in duplicate, and five MALDI spectra obtained from each for a total of 10 measurements per reaction. Product yields were calculated from the average ratio of observed product and template ion counts using a standard curve generated from independently quantitated and synthesized authentic product (see Supporting Information).<sup>11</sup>

Addition of a single base to an abasic site at the middle of the PNA strand using reductive amination proceeds with good to excellent yield (78–98%) and good selectivity ( $\geq 7:1$  ratio of the matched product to each possible mismatch product) for each of the four bases (Table 2, entries a–d). Small amounts of mismatched products are observed in the additions of adenine and thymine, but

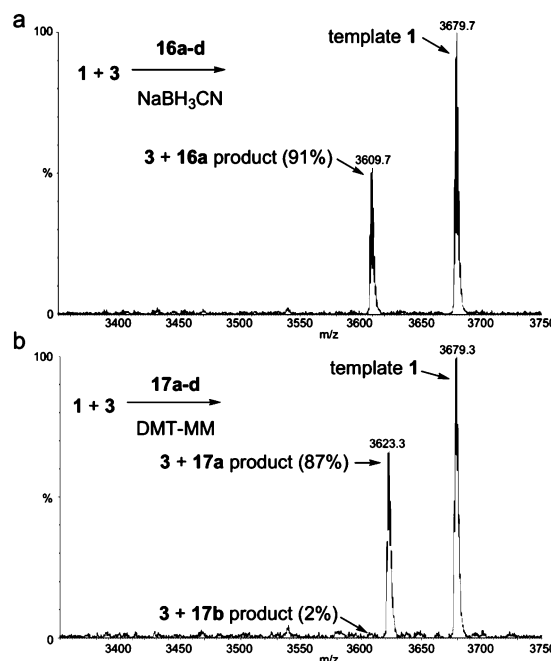
**Figure 3.** Nucleobase substrates for base-filling reactions.

the impurities in the adenine addition can be attributed to formation of a T:G wobble pair.

For middle-of-strand base filling by amine acylation (Table 2, entries a–d), we observed a much broader range of results than for reductive amination depending on the identity of the substrate base and the acylation reagents used. Using EDC and sNHS, we observed good selectivity for all four reactions but good yield (66–78%) only when adding guanine or cytosine. Adenine addition proceeds with a significantly lower yield (38%), and the yield for thymine addition is poor (8%). Switching the acylation reagent to DMT-MM increases the yield for purine addition but lowers the yields and the sequence specificity for pyrimidine additions.

These variations in acylation efficiency and selectivity likely arise from the different steric and  $\pi$ -stacking characteristics of the activated ester intermediates. Our results also suggest that, for middle-of-strand base filling, G:C base pairing supports more efficient product formation than A:T base pairing. Further, despite producing canonical PNA, acylation is generally less efficient and selective than reductive amination. A similar trend has been observed for other DNA-templated reaction substrates<sup>1f</sup> and is likely a consequence of the reversible formation of an iminium intermediate during reductive amination.<sup>1d</sup>

We next investigated base filling of a single abasic site at the end of the PNA strand (Table 2, entries e–h). These reactions are more challenging, since the base-stacking interaction between the incoming base and the template should be significantly weaker than the case in the middle of a strand. Under reductive amination conditions, we observed moderate yields for the additions of guanine and adenine (47–58%) and low yields for the additions of cytosine and thymine (2–7%). Under acylation conditions, we observed no significant addition of any of the four bases using EDC/sNHS, but DMT-MM provided moderate yields (39–42%) for the additions of guanine and adenine and poor yields (4–28%) for the additions of cytosine and thymine. The selectivities of all of the reactions were significantly lower than those observed for the middle-of-strand additions. The universally lower yields for the end-of-strand reactions in Table 2 implicate base-stacking interactions and the

**Figure 4.** MALDI spectra of sequence-specific base filling of a guanine nucleobase (**16a** or **17a**) at the abasic site in **3** via (a) reductive amination and (b) amine acylation. See Table 2 for reaction conditions.

**Table 2.** Yields of base-filling reactions<sup>a</sup>

entry	PNA duplex <sup>b</sup>	matched base	reductive amination		acylation	
			yield (%), matched	yield (%), mismatched	yield (%), matched	yield (%), mismatched
a	1 <sup>N</sup> TAC TGC TAG ACG <sup>C</sup> 3 <sup>C</sup> ATG AC_ ATC TGC <sup>N</sup>	G	91 <sup>c</sup>	<2 <sup>c</sup>	78 <sup>d</sup> 87 <sup>e</sup>	<2 <sup>d</sup> 2(A) <sup>e</sup>
b	1 <sup>N</sup> TAC TGC TAG ACG <sup>C</sup> 4 <sup>C</sup> ATG ACG _TC TGC <sup>N</sup>	A	78 <sup>c</sup>	6(G) <sup>c</sup>	38 <sup>d</sup> 73 <sup>e</sup>	5(G) <sup>d</sup> 7(G) <sup>e</sup>
c	2 <sup>N</sup> CGT CTA GCA GTA <sup>C</sup> 5 <sup>C</sup> GCA GA_ CGT CAT <sup>N</sup>	T	81 <sup>c</sup>	11(G), 4(C) <sup>c</sup>	8 <sup>d</sup> <2 <sup>e</sup>	<2 <sup>d</sup> 19(A), 13(G), 5(C) <sup>e</sup>
d	2 <sup>N</sup> CGT CTA GCA GTA <sup>C</sup> 6 <sup>C</sup> GCA GAT _GT CAT <sup>N</sup>	C	98 <sup>c</sup>	<2 <sup>c</sup>	66 <sup>d</sup> 16 <sup>e</sup>	<2 <sup>d</sup> 26(G), 17(A) <sup>e</sup>
e	2 <sup>N</sup> CGT CTA GCA GTA <sup>C</sup> 7 <sup>C</sup> _CA GAT CGT CAT <sup>N</sup>	G	58 <sup>c</sup>	13(A), 9(C) <sup>c</sup>	39 <sup>e</sup>	16(A), 5(C) <sup>e</sup>
f	1 <sup>N</sup> TAC TGC TAG ACG <sup>C</sup> 8 <sup>C</sup> _TG ACG ATC TGC <sup>N</sup>	A	47 <sup>c</sup>	29(G), 11(T), 8(C) <sup>c</sup>	42 <sup>e</sup>	26(G), 7(C) <sup>e</sup>
g	2 <sup>N</sup> CGT CTA GCA GTA <sup>C</sup> 9 <sup>C</sup> GCA GAT CGT CA_ <sup>N</sup>	T	2 <sup>c</sup>	4(C), 3(G), 2(A) <sup>c</sup>	4 <sup>e</sup>	15(C), 7(G), 6(A) <sup>e</sup>
h	1 <sup>N</sup> TAC TGC TAG ACG <sup>C</sup> 10 <sup>C</sup> ATG ACG ATC TG_ <sup>N</sup>	C	7 <sup>c</sup>	4(G), 4(A), 3(T) <sup>c</sup>	28 <sup>e</sup>	10(A), 9(G), 9(T) <sup>e</sup>
i	1 <sup>N</sup> TAC TGC TAG ACG <sup>C</sup> 13 <sup>C</sup> ATG AC_ _TC TGC <sup>N</sup>	G–A	38 <sup>c</sup>	7(G–G), 3(G–C), 3(C–A) <sup>c</sup>	56 <sup>e</sup>	23(G–G), 5(G–C), 4(C–A), 3(A–A) <sup>c</sup>
j	14 <sup>N</sup> TGT ACG CGA TCA <sup>C</sup> 15 <sup>C</sup> ACA TG_ _CT AGT <sup>N</sup>	C–G	21 <sup>c</sup>	6(G–G), 4(A–G) <sup>c</sup>	9 <sup>e</sup>	16(G–G), 12(A–G), 3(T–G) <sup>e</sup>

<sup>a</sup> Yields shown each represent the mean of 10 measurements from two independent reactions and are accurate to  $\pm 10\%$ . <sup>b</sup> Lys(Me<sub>2</sub>) and capping groups are omitted for clarity. <sup>c</sup> 3  $\mu\text{M}$  template, 2.5  $\mu\text{M}$  reactant strand, 150  $\mu\text{M}$  each **16a–d**, 15 mM NaBH<sub>3</sub>CN, 10 mM phosphate buffer, pH 6.2, 5 °C, 24 h. <sup>d</sup> 3  $\mu\text{M}$  template, 2.5  $\mu\text{M}$  reactant strand, 830  $\mu\text{M}$  each **17a–d**, 9 mM sNHS, 13 mM EDC, 100 mM MOPS buffer, pH 7.5, 5 °C, 24 h. <sup>e</sup> 3  $\mu\text{M}$  template, 2.5  $\mu\text{M}$  reactant strand, 830  $\mu\text{M}$  each **17a–d**, 13 mM DMT-MM, 100 mM MOPS buffer, pH 7.5, 2 °C, 24 h.

lower disorder of double helices away from strand termini as key determinants of base-filling reaction efficiency.

To explore the ability of base filling to support the transfer of multiple consecutive bases, we attempted the addition of two neighboring bases to the middle of the PNA strand (Table 2, entries i–j). We used PNA template **1** to test the tandem addition of guanine and adenine and introduced a new template (**14**) to test the tandem addition of cytosine and guanine. We observed significant yields of doubly base-filled product for both reductive amination (up to 38%) and amine acylation (up to 56%), although yields were lower than the analogous single-base addition reactions. Consistent with our observations above, purines added more efficiently than pyrimidines (comparing the addition of adenine+guanine with that of cytosine+guanine), and the reductive amination reactions proceeded with higher selectivities than the acylation reactions.

In conclusion, we have achieved sequence-selective templated nucleic acid synthesis through base filling. Using either reductive amination or acylation reactions, a single base can be sequence-specifically added to an abasic site on a PNA strand in the presence of all four nucleobase monomers in the reaction mixture. We observe that middle-of-strand addition is more efficient than end-of-strand addition and that purines add more efficiently than pyrimidines, highlighting the role of base-stacking interactions in promoting base-filling reactions. Base-pairing interactions may also play a significant role, as in general guanine adds more efficiently than adenine and cytosine adds more efficiently than thymine. Additionally, we observe that reductive amination generally provides higher yields and selectivities than the amine acylation chemistries tested, likely resulting from the reversible formation of a stable iminium intermediate in the former reaction. These findings lay an experimental foundation for the development of more complex information transfer systems, including the possibility of exponential replication of nucleic acids through base-filling reactions.

**Acknowledgment.** This work was supported by the Howard Hughes Medical Institute and the NIH/NIGMS (R01GM065865).

**Supporting Information Available:** Experimental procedures and compound characterization data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) (a) Naylor, R.; Gilham, P. T. *Biochemistry* **1966**, *5*, 2722–2728. (b) Inoue, T.; Joyce, G. F.; Grzeskowiak, D.; Orgel, L. E.; Brown, J. M.; Reese, C. B. *J. Mol. Biol.* **1984**, *178*, 669–676. (c) von Kiedrowski, G. *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 932–935. (d) Li, X.; Zhan, Z.-Y.; Knipe, R.; Lynn, D. G. *J. Am. Chem. Soc.* **2002**, *124*, 746–747. (e) Xu, Y.; Karalkar, N. B.; Kool, E. T. *Nat. Biotechnol.* **2001**, *19*, 148–152. (f) Rosenbaum, D. M.; Liu, D. R. *J. Am. Chem. Soc.* **2003**, *125*, 13924–13925. (g) Chen, J. J.; Cai, X.; Szostak, J. W. *J. Am. Chem. Soc.* **2009**, *131*, 2119–2121.
- (2) For a theoretical description of base filling, see: Hickman, D. T.; Sreenivasachary, N.; Lehn, J.-M. *Helv. Chim. Acta* **1998**, *91*, 1–20.
- (3) While this manuscript was under review, base filling via thioester bond formation was reported; see: Ura, Y.; Beierle, J. M.; Leman, L. J.; Orgel, L. E.; Ghadiri, M. R. *Science* **2009**, *325*, 73–77.
- (4) Orgel, L. E.; Lohrmann, R. *Acc. Chem. Res.* **1974**, *7*, 368–377.
- (5) N-Glycosidic bond formation is easily achieved in organic solvents. For an example, see: Schweitzer, B. A.; Kool, E. T. *J. Org. Chem.* **1994**, *59*, 7238–7242.
- (6) Unrau, P. J.; Bartel, D. P. *Nature* **1998**, *395*, 260–263.
- (7) Nielsen, P. E. *Chem. Biodiversity* **2007**, *4*, 1996–2002.
- (8) Hyrup, B.; Egholm, M.; Buchardt, O.; Nielsen, P. E. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1083–1088.
- (9) Shukla, V.; Mishra, S.; Watal, G.; Misra, K. *Ind. J. Chem. B* **2005**, *44B*, 121–129.
- (10) Monomers **16a–d** are depicted as aldehydes in Figure 3 for clarity, but under the aqueous reaction conditions used, they are observed to exist predominantly in the hydrated form.
- (11) The use of a standard curve based on a product/template ion count ratio normalizes potential differences in relative peak height arising from either variable retention during sample desalting or differences in MALDI ionization efficiency.

JA904712T