



# HHS Public Access

Author manuscript

*Nat Chem Biol.* Author manuscript; available in PMC 2010 August 01.

Published in final edited form as:

*Nat Chem Biol.* 2010 February ; 6(2): 148–155. doi:10.1038/nchembio.280.

## An *In Vitro* Translation, Selection, and Amplification System for Peptide Nucleic Acids

Yevgeny Brudno, Michael E. Birnbaum, Ralph E. Kleiner, and David R. Liu\*

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

### Abstract

Methods to evolve synthetic, rather than biological, polymers could significantly expand the functional potential of polymers that emerge from *in vitro* evolution. Requirements for synthetic polymer evolution include: (i) sequence-specific polymerization of synthetic building blocks on an amplifiable template; (ii) display of the newly translated polymer strand in a manner that allows it to adopt folded structures; (iii) selection of synthetic polymer libraries for desired binding or catalytic properties; and (iv) amplification of template sequences surviving selection in a manner that allows subsequent translation. Here we report the development of such a system for peptide nucleic acids (PNAs) using a set of twelve PNA pentamer building blocks. We validated the system by performing six iterated cycles of translation, selection, and amplification on a library of  $4.3 \times 10^8$  PNA-encoding DNA templates and observed >1,000,000-fold overall enrichment of a template encoding a biotinylated (streptavidin-binding) PNA. These results collectively provide an experimental foundation for PNA evolution in the laboratory.

---

The dominance of nucleic acids and proteins among biological molecules with complex functional properties is likely a consequence of their ability to evolve. During biological evolution (Fig. 1a), a protein or nucleic acid with properties that enhance an organism's fitness increases the probability that the gene encoding its structure survives. This genetic information is replicated through the division of a cell or organism, gradually diversified through mutation and recombination, and translated into related biological polymers that determine the ability of a new generation of cells or organisms to survive.

In the laboratory, directed evolution has historically been limited to two types of polymers—proteins and nucleic acids—because for decades they were the only polymers that could be translated from the information in DNA. The prospect of evolving polymers with functional properties that might extend beyond those accessible by DNA, RNA, and proteins has driven the desire to apply molecular evolution to synthetic, rather than biological, polymers<sup>1,2</sup>. A

---

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\*drliu@fas.harvard.edu.

#### **Author Contributions**

Y.B. and D.R.L. designed the research. Y.B., M.E.B., and R.E.K. performed the experiments. All authors contributed to data analysis and manuscript preparation.

#### **Competing Financial Interests**

Y.B. and D.R.L. are co-inventors on a Harvard University patent describing DNA-templated polymerization.

Author Manuscript

general approach to this goal (Fig. 1b) requires the sequence-specific translation of nucleic acid templates into synthetic polymers, the selection of libraries of synthetic polymers generated in this manner, and the amplification of nucleic acids that encode selection survivors in a manner that enables a subsequent round of translation. In principle, such an approach could enable the directed evolution of sequence-defined synthetic polymers with tailor-made functional properties.

Author Manuscript

One strategy used to accomplish the crucial translation step exploits the ability of the ribosome or polymerase enzymes to accept synthetic variants of aminoacylated tRNAs or nucleotide triphosphates, respectively. While these approaches have proven successful<sup>3–5</sup>, the required compatibility with the biosynthetic enzymes imposes major structural constraints. As a result, synthetic polymers of significant length translated in this manner typically conform to a modest set of modifications that are accepted by the protein or nucleic acid biosynthetic machinery.

Author Manuscript

Non-enzymatic, nucleic acid-templated polymerization represents a different strategy to synthetic polymer translation<sup>6,7</sup>. We previously reported the efficient and sequence-specific translation of DNA templates into peptide nucleic acids (PNAs) using DNA-templated reductive amination as the building-block coupling reaction<sup>8,9</sup>. PNA<sup>10</sup> binds to DNA templates sequence-specifically via Watson-Crick base-pairing, can be readily synthesized in forms that contain a variety of side-chain functional groups<sup>8,11</sup> and in its single-stranded form has been observed to adopt three-dimensional structures beyond canonical double helices<sup>12–15</sup>. PNA aptamers and catalysts may complement existing nucleic acid aptamers and protein antibodies by offering improved cellular or chemical stability, lower immunogenicity<sup>16</sup>, improved cell penetration<sup>17</sup>, or resistance to degradation<sup>18</sup>.

Author Manuscript

Several key challenges must be met to evolve PNAs using non-enzymatic polymerization. A population of DNA templates must direct the efficient, high-fidelity polymerization of PNA building blocks into corresponding PNA oligomers using a method that is general for many different sequences. Since two-stranded DNA-PNA heteroduplexes likely adopt only a limited range of double helical structures, the newly polymerized PNA strand must be efficiently displaced from the DNA template to allow the PNA to spontaneously fold into secondary and tertiary structures. The folded PNA can then be subjected to an *in vitro* selection for desirable properties such as target affinity or the ability to catalyze a chemical reaction. The DNA sequences corresponding to PNAs surviving the selection must be amplified, and occasionally mutated, in a manner that recreates DNA templates suitable for a subsequent round of translation, displacement, selection, and amplification. Throughout the translation and selection steps, a stable linkage associating each DNA sequence with its corresponding PNA must be maintained.

Author Manuscript

In this work we report a selection system for the directed evolution of PNAs that meets these requirements. We describe a systematic study of DNA-templated polymerization reactions of PNA tetramer and pentamer building blocks that led to an optimized PNA “genetic code.” We demonstrate the successful translation of a variety of DNA templates into PNA oligomers containing a broad mixture of PNA pentamer building blocks from our optimized genetic code. We also describe an efficient method of PNA strand displacement used to

disrupt base-pairing between PNA products and DNA templates while preserving the PNA-DNA covalent linkage. Finally, we demonstrate >1,000,000-fold enrichment of a DNA template encoding a unique biotinylated library member from a large PNA library through iterated cycles of translation, displacement, selection for streptavidin binding, amplification of the selection survivors, and conversion of the amplified DNA sequence into new templates that are capable of undergoing subsequent translation into PNA. Collectively, these developments represent the first selection system based on non-enzymatic polymerization and make possible the directed evolution of a sequence-defined synthetic polymer that is not compatible with biosynthetic machinery.

## Results

### Development of a PNA Genetic Code: The Challenge of Uniform Polymerization

The “genetic code” of a PNA selection system defines the set of PNA building blocks used during translation and should meet at least four criteria. First, all building blocks in the code should polymerize with comparable and high efficiency under a given set of conditions to minimize the fraction of the library that is lost to unsuccessful translation. Second, the code should preclude significant hybridization between any DNA codon and a non-cognate (mismatched) PNA building block to avoid mistranslation. Third, the genetic code should minimize the possibility of a building block polymerizing out of frame. Finally, the DNA templates encoding a library of PNA polymers should be readily prepared in a single solid-phase DNA synthesis.

Our previous studies established the ability of DNA-templated reductive amination reactions to mediate the polymerization of PNA aldehyde oligomers on DNA hairpin templates<sup>8,9</sup> (Fig. 1c). Although these reactions exhibited promising degrees of efficiency and sequence specificity when applied to DNA templates encoding predominantly gact or ggatt building blocks (where lower-case letters represent PNA nucleotides), they have not previously been applied to the more complex problem of translating templates containing many different codons or translating libraries containing a mixture of many different templates simultaneously.

We previously observed that polymerization efficiency correlates well with the melting temperature ( $T_m$ ) of individual building blocks hybridized to the DNA template<sup>8</sup>. We therefore sought models that could predict the  $T_m$  of PNA building blocks and by extension their polymerization potential. Because methods for predicting DNA-PNA duplex melting temperatures differ<sup>19,20</sup>, particularly in their handling of the pyrimidine content (the fraction of pyrimidines) in a sequence, we first determined which factors accurately predict the efficiency of DNA-templated PNA aldehyde polymerization.

We began by studying the polymerization of eight PNA tetramer building blocks used in our earlier studies<sup>9</sup>. In contrast with our earlier studies in which we examined polymerization on a template of fixed flanking sequences with a single centrally located variable codon,<sup>9</sup> here we sought to reveal subtle differences in polymerization efficiencies by using DNA templates consisting of ten consecutive repeats of a particular codon (Supplementary Fig. 1a). The templated polymerization of the eight building blocks on repetitive codon templates

at 25 °C, 45 °C, and 60 °C is shown in Supplementary Fig. 1b. We observed dramatic differences in polymerization efficiency among these building blocks when studied in the repetitive codon format. One building block, gcct, with a pyrimidine content of 75%, did not polymerize at room temperature. Four building blocks, gaat, gact, gcgt and ggct (50% or 25% pyrimidine), were able to polymerize at 45 °C, but not at 60 °C. The other three building blocks, gagt, ggat and gggt (all 25% pyrimidine), were able to polymerize to some extent at 60 °C. These results indicate that pyrimidine content has a strong impact on DNA-templated PNA polymerization efficiency, consistent with the model proposed by Nielsen and coworkers<sup>19</sup>. In addition, these results also suggest that GC content plays a role in polymerization efficiency: although gggt, gagt, ggat and gaat all have identical pyrimidine content, the first three are able to polymerize at the highest temperature while the final one fails to do so, presumably due to its lower GC content.

To better assess the effect of pyrimidine content on PNA building block polymerization, we studied the polymerization of ten pentamer PNA building blocks of identical GC content. Polymerizations were performed on DNA templates consisting of seven consecutive repeats of a particular codon. The ten building blocks each contained two g or c nucleotides (*i.e.*, g-C or c-G base pairs) and three a or t nucleotides (*i.e.*, a-T or t-A base pairs), but varied in PNA pyrimidine content (c or t) from 20% to 80%. Supplementary Fig. 1c shows the polymerization of the ten building blocks at 25 °C, 45 °C, and 60 °C. We again observed large differences in polymerization efficiency between building blocks, despite their identical GC content. The two building blocks with lowest pyrimidine content (aaggt and agagt) were able to polymerize efficiently even at 60 °C, while three others (cttgt, ccatt and ctact) failed to polymerize even at 25 °C. In agreement with the Nielsen formula and other published reports<sup>21</sup>, we again observed that lower pyrimidine content correlated with higher polymerization efficiency.

Taken together, these results (Supplementary Results) are consistent with the Nielsen model for PNA-DNA hybridization strength and suggest that uniform polymerization behavior at a given temperature, a key criterion for a PNA selection system, is optimal when all building blocks in the genetic code have identical GC and pyrimidine content.

### **Development of a PNA Genetic Code: Ease of Template Library Synthesis, Sequence Specificity, and Uniform Polymerization**

We sought to integrate the requirements for uniform GC and pyrimidine content elucidated above with three other requirements necessary for a viable PNA genetic code: high sequence specificity, in-frame hybridization to a template, and ease of library synthesis. Based on our previous studies<sup>9</sup> we hypothesized that ensuring at least one non-G:T mismatch or at least two G:T wobble mismatches between any DNA codon and any non-matched PNA building block would be sufficient to support sequence-specific polymerization for short building blocks. The genetic code should also minimize the possibility that building blocks polymerize out of frame to avoid mistranslation. Finally, an ideal PNA genetic code enables DNA template libraries to be readily synthesized in a single column on a DNA synthesizer without the need to perform laborious split-and-pool oligonucleotide synthesis.

We considered all possible sets of PNA building blocks one to four nucleotides in length and concluded that these lengths do not offer a genetic code of sufficient size and complexity for a PNA selection system (see the Supplementary Methods for a detailed analysis). However, we identified a 12-building block set of pentanucleotide PNAs (acact, accat, acgtt, caact, cacat, cagtt, tgact, tgcac, tgggt, gtact, gtcac, gtggt) that satisfy all of the above constraints. The DNA template library can be synthesized in a single column by using a coding region of repeating AY'X' codons, where Y' represents a degenerate mixture of AC, GT, and TG dinucleotide phosphoramidites and X' represents a degenerate mixture of CA, AC, GT, and TG dinucleotide phosphoramidites. We proceeded to characterize the properties of this PNA genetic code (x'y't) in depth.

To assay the uniformity of polymerization of the 12-building block set, we performed DNA-templated polymerizations at 25 °C, 45 °C, and 60 °C (Fig. 2). Eight of the building blocks were tested using repetitive codon templates containing ten consecutive repeats of the corresponding codon. For four of the building blocks (acgtt, gtact, tgcac, and gtcac), the repetitive codon templates adopt internal secondary structure. Consequently, these building blocks were tested using templates in which the corresponding codons were alternated with the AACCA codon. Importantly, all members of the genetic code polymerize with comparable efficiencies (including no observed polymerization at the highest temperature tested), in sharp contrast with the variation observed within the earlier tetramer and pentamer PNA genetic codes (Supplementary Fig. 1). These results establish the ability of the above 12-building block PNA pentamer genetic code to polymerize uniformly across a range of temperatures and validate several of the design principles described above.

Our previous experiments to probe the sequence specificity of DNA-templated PNA polymerization examined the incorporation of building blocks on a template of fixed flanking sequences with a single centrally located variable codon<sup>9</sup>. To more stringently ascertain the sequence specificity of polymerization, we studied polymerization specificity at each position of a six-codon (30-base) template, and in the context of a wide variety of neighboring codons. We prepared four templates that collectively used each of the 12 codons twice in a random order. We prepared all twelve corresponding PNA aldehyde pentamer building blocks in both a free amine form (NH<sub>2</sub>-x'y't-CHO) and a capped form in which the amino terminus of the building block was acetylated (AcNH-x'y't-CHO). The capped building blocks can still react via their C-terminal aldehyde group, but terminate polymerization because they cannot support subsequent coupling reactions.

We performed DNA-templated polymerizations using each of these templates and mixtures containing 11 free amine building blocks and one capped building block (Fig. 3). When a capped building block is correctly incorporated in the middle of a strand, polymerization terminates; if an incorrect building block is incorporated, however, the polymerization reaction can read through the codon for the acetylated monomer, resulting in full-length polymer. For each of the 12 building blocks in all four mixed-sequence templates, we observed sequence-specific incorporation of the capped, matched building block at the correct position in the template to generate predominantly a truncated product of expected length, even in the presence of an 11-fold combined excess of the other mismatched, non-acetylated building blocks (Fig. 3). These results establish that the twelve building blocks

are incorporated in a sequence-dependent manner with little misincorporation of mismatched building blocks. Importantly, we observed similar results for all 24 codons among the four mixed-sequence templates, indicating that polymerization remained sequence-specific even when the codons preceding and following the codon being tested varied widely.

Taken together, these experiments reveal a 12-building block PNA pentamer genetic code that exhibits uniform reactivity at a range of temperatures, consists entirely of codons that can be accessed in a single-column DNA template library synthesis using degenerate dinucleotide phosphoramidites, and maintains efficiency and sequence specificity even in the presence of a mixture of different PNA building blocks and in a variety of sequence contexts. In light of these properties, we used the x'y't PNA genetic code throughout the studies described below.

### Displacement of Translated PNA Strands from DNA Templates

Before a translated PNA can undergo *in vitro* selection for functional properties, it must be liberated from the DNA template to which it is hybridized so that it can adopt a native three-dimensional conformation. Work by Szostak and coworkers has demonstrated that polymerases can efficiently displace threose nucleic acid (TNA) oligomers<sup>22</sup>. PNA, however, is structurally distinct from sugar-phosphate-based nucleic acids such as TNA, DNA, or RNA and it was unclear if polymerase-mediated strand displacement would be possible for DNA-PNA duplexes. Indeed, DNA-PNA duplexes are typically more stable than their corresponding DNA-DNA duplexes<sup>10</sup>, and can invade DNA-DNA duplexes<sup>23-25</sup>, increasing the challenge of using DNA primer extension to displace the PNA strand of a DNA-PNA heteroduplex.

We speculated that seemingly unfavorable PNA strand displacement could be achieved by providing a thermodynamic incentive of additional DNA-DNA base pairs that are only formed upon strand displacement. We therefore studied PNA displacement using a template architecture containing 3' and 5' stem-loops (Fig. 4). After PNA polymerization, the PNA strand is displaced by primer extension of the 3' stem-loop terminus. Successful primer extension to the end of the template strand (into the 5' stem-loop) generates a 100-base pair DNA duplex that can form as a mutually exclusive alternative to the 40-base pair DNA-PNA heteroduplex connected to a 25-base pair DNA-DNA duplex (Fig. 4). We hypothesized that the significant gain in the number of DNA-DNA base pairs could overcome the superior stability of the DNA-PNA duplex. Because this approach generates double-stranded DNA templates, it also helps prevent the DNA component of the resulting library from folding into conformations that may influence the outcome of a selection.

We tested a panel of mesophilic and thermophilic polymerases for their ability to displace an eight-building block (40-nucleotide) PNA oligomer from a PNA-DNA heteroduplex in the double hairpin template architecture. Enzymes that successfully displace the PNA create a DNA-DNA duplex containing a digestion site for the restriction enzyme *BccI*, whereas the starting DNA-PNA duplex cannot be digested by *BccI* (Fig. 4). Herculanase II DNA polymerase (Stratagene), a fusion protein of Pfu Ultra and a DNA-binding domain that is designed to facilitate DNA polymerization on GC-rich templates, was able to displace the

PNA strand efficiently (Fig. 4). The nearly quantitative efficiency of the restriction digestion of Herculase II products suggests that virtually all the product molecules exist in a form in which the translated PNA strand is no longer base-paired with the DNA template, consistent with successful PNA strand displacement (Fig. 4). We also identified a length limit to the efficient displacement of PNA from DNA. When Herculase II was challenged with the displacement of a 10-building block (50-bases) DNA-PNA heteroduplex, displacement efficiency was significantly lower.

The most stably folded single-stranded PNAs likely represent the library members of greatest functional potential. We anticipate that these highly folded PNAs will be the least likely library members to revert to a PNA-DNA hybridized form because their PNA folding energy, in addition to the extra DNA-DNA base pairs created upon PNA strand displacement, contributes to the stability of the single-stranded PNA form. To gain insight into the potential of our library to include well-folded members, we used the oligonucleotide modeling platform (OMP)<sup>26</sup> to simulate the intramolecular secondary structure of 10,000 randomly chosen (x'y't)<sub>8</sub> library members compared with 10,000 randomly chosen N<sub>40</sub> library members. Since rapid PNA structure modeling was not available, both libraries were modeled as DNA. The results suggests that the (x'y't)<sub>8</sub> library and the N<sub>40</sub> library exhibit similar folding energies and similar distributions ( $\Delta G = -3.6 \pm 2.1$  kcal/mol for N<sub>40</sub> versus  $-3.7 \pm 1.9$  kcal/mol for the (x'y't)<sub>8</sub> library) (Supplementary Fig. 2).

### **Amplification and Manipulation of DNA Templates From Displaced PNA-dsDNA Conjugates**

Iterated cycles of selection and amplification<sup>27</sup> allow even extremely rare but highly functional library members to become well represented. In order for PNAs to undergo multiple iterated selection cycles, DNA templates encoding PNA molecules that survive a selection must be amplified and manipulated in a manner that installs the 5' and 3' hairpins needed for a subsequent round of translation. We used a three-step procedure to achieve these goals (Fig. 5 and Supplemental Methods). First, the coding regions of templates surviving selection were amplified in a PCR reaction using two modified primers. Second, from the resulting double-stranded PCR product (Fig. 5, lane 1) the desired coding strand was separated from the non-coding strand and isolated. Finally, the resulting single-stranded DNA (Fig. 5, lane 2) was ligated to two DNA hairpins in a one-pot reaction catalyzed by T4 DNA ligase. Both ligation reactions are self-templating and proceed very efficiently (Fig. 5, lanes 3, 4 and 5).

To demonstrate that these steps enable a complete cycle of translation, displacement, and amplification to take place, we subjected the resulting coding strand flanked by 5' and 3' hairpins to DNA-templated PNA polymerization (Fig. 5, lane 6), followed by PNA strand displacement (Fig. 5, lane 7) as described above. To complete the cycle, a second PCR amplification was performed on 1/15,000th of the translated and displaced material to simulate a post-selection amplification reaction (Fig. 5, lane 8). The resulting PCR product was indistinguishable by gel electrophoresis from the PCR product used to initiate the cycle (Fig. 5, lane 1). Taken together, these results validate the component steps developed in the

studies above and collectively represent a complete cycle of translation, displacement, and amplification of a DNA sequence encoding a sequence-defined synthetic polymer.

### Iterated Translation and Selection of a Biotinylated PNA from a Model Library

To test the ability of the PNA selection system to support translation, displacement, selection, and amplification in a library format, we performed these manipulations on a model DNA-templated PNA library. We synthesized a library of DNA templates containing eight consecutive five-base codons in a single solid-phase synthesis using degenerate mixtures of dinucleotide phosphoramidites. Each of the eight coding positions contained an equimolar mixture of 12 codons corresponding to the x'y't PNA genetic code; thus, the template library coding region was (AY'X')<sub>8</sub> where Y' is an equimolar mix of AC, GT, and TG, and X' is an equimolar mix of CA, AC, GT, and TG, representing a total theoretical library complexity of  $4.3 \times 10^8$ . Cloning and DNA sequencing of 50 templates (400 codons) from the library revealed a sequence composition of 33% AC, 35% GT, and 32% TG at the Y' position and 21% CA, 24% AC, 25% GT, and 30% TG at the X' position, in good agreement with the template library design. Selections of DNA-templated libraries are commonly performed on ~1 pmol ( $6 \times 10^{11}$  molecules) of library material.<sup>28,29</sup> We calculate that every member of our  $4.3 \times 10^8$ -membered library has a greater than 99.9% chance of being present in each selection performed on a pmol scale.

Next we performed a mock selection that tested the behavior of DNA templates and PNA polymers over multiple iterated cycles of translation, displacement, selection, and amplification. In a single solution we combined the DNA template library with 1/100th, 1/10,000th or 1/1,000,000th of one equivalent of a positive control DNA template that uniquely contains the coding sequence (AGTGT)<sub>7</sub>AATCC and an *MspI* restriction enzyme cleavage site. The final AATCC codon was not present in any other library members and encodes a biotinylated ggatt PNA building block. Templates were translated into PNA oligomers, bound to streptavidin beads, and eluted by digestion with *AvaI*, which cleaves DNA templates from protein-bound PNAs. This experimental design serves as a test of the sequence specificity of PNA polymerization and the ability of a single library member to be enriched by iterated *in vitro* selection for protein affinity. The biotinylated ggatt building block should only be incorporated into PNA polymers encoded by the AATCC-containing template, and multiple selection cycles should eventually enrich this template even if highly underrepresented in the starting library.

Six complete rounds of selection for streptavidin binding, amplification by PCR, and retranslation into PNA were performed. To ensure that any observed enrichment was the result of selection for streptavidin binding, rather than an artifact of the iterated selection protocol, a 1:100 mixture of positive control template to library was also submitted to six rounds of selection in which each PNA translation step lacked the biotinylated building block.

The extent to which the translation, displacement, selection, and amplification process enriched the positive control template was evaluated by restriction digestion with *MspI*. After two rounds of translation, selection, and amplification, the 1:100 starting mixture of positive control: other library members become mostly positive control templates. For the



1:10,000 starting mixture, the positive control template becomes the major species in the library after three rounds of selection (Supplementary Fig. 3). The library beginning with a 1:1,000,000 mixture of positive control to other library members was transformed into a mixture containing mostly positive control templates after six iterated rounds of selection (Fig. 6). These results are consistent with the range of enrichment levels previously reported for *in vitro* selections of DNA-linked protein ligands<sup>28,30,31</sup>.

No DNA was observed to elute from samples that underwent translation but not PNA strand displacement, suggesting that the observed enrichment was dependent on strand displacement of a PNA containing the correctly incorporated biotinylated building block. Likewise, no enrichment was observed in the 1:100 control case in which six rounds of selection were performed with the biotinylated building block omitted from each PNA translation (Fig. 6), indicating that the amplification observed after six rounds is not an artifact of repeated DNA manipulation, but instead depends on the incorporation of the biotinylated building block. These results establish the ability of the methods developed in this work to significantly enrich a single, rare DNA template from a complex library based on the protein-binding properties of an encoded biotinylated building block. Importantly, these findings also demonstrate the feasibility of iterating translation, selection, and amplification over multiple rounds to multiply the net enrichment of DNA sequences encoding active PNA library members<sup>28,32,33</sup>, a crucial feature of existing biological polymer evolution systems.

## Discussion

In this work we have designed and implemented a complete system for the *in vitro* translation, selection, and amplification of DNA sequences encoding a synthetic polymer. In the process, we elucidated key properties of PNA building blocks needed to achieve uniform polymerization and sequence specificity, and identified robust conditions for the displacement of a PNA strand from a DNA-PNA duplex. Integrating these developments, we performed multiple iterated cycles of translation, selection, and amplification of DNA sequences encoding sequence-defined PNA oligomers and observed the enrichment of a single DNA template, which required both sequence-specific PNA translation and PNA strand displacement.

The neutrality and hydrogen bonding potential of the PNA backbone make PNA a unique candidate for the discovery of functional oligomers. The function of biological polymers typically requires a well-defined three-dimensional fold, the formation of which can be hindered if the polymer backbone is excessively charged because charge-charge repulsion in the folded state must be neutralized or offset by other enthalpic and entropic factors<sup>34</sup>. Polymers with uncharged or weakly charged backbones such as the PNAs generated in this work therefore may be able to adopt a greater diversity of folded conformations. The lower charge density of PNA compared with DNA, RNA, and other phosphate backbone-based nucleic acid analogs such as TNA and LNA may increase PNA's propensity to adopt three-dimensional folds and facilitate the evolution of functional receptors and catalysts.

The polymer backbone generated in this work contains secondary amine groups every fifth PNA nucleotide, a property that may both benefit and impair the selection of functional PNAs. Studies on single-stranded PNAs suggest that PNA lacking explicitly designed secondary structure frequently exist as compact globules<sup>35,36</sup>, similar to random sequences of amino acids<sup>37</sup>. The additional charge present in the PNAs generated in our system may help to counter the formation of aggregates in solution in favor of more productively folded, monomeric PNAs. Alternatively, the moderate degree of cationic character may induce some degree of non-specific electrostatic association with DNA templates, compromising folding and selection. How the structural details of any synthetic polymer affect its ability to be selected or evolved represents a fascinating and important question that can be studied using systems such as the one described in this work. In addition, the evolution of synthetic polymers with certain properties may require the incorporation of other functionality into the backbone, including negatively charged or hydrophilic groups, a possibility we discuss below.

The length and complexity of a synthetic polymer library are important determinants of the likelihood that the library contains one or more members with desired functional properties. The theoretical complexity of the PNA library described in this work,  $4.3 \times 10^8$  polymers, lies within the range of  $10^6$ - to  $10^{10}$ -membered peptide libraries typically generated by phage display or cell-surface display, but is small in comparison to traditional RNA and DNA libraries that can consist of  $10^{12}$ – $10^{15}$  library members<sup>32,33,38,39</sup>. Because the degree to which PNA sequence space is populated with islands of functional sequences is unknown, it is difficult to predict how many functional sequences might occur in the PNA library described here when challenged with a particular binding or catalytic task. It is possible that eight repeats of pentanucleotide building blocks will offer too little complexity to evolve PNAs with desired functions. If so, alternative strategies can be considered that enable the displacement of more pentanucleotide building blocks per PNA strand or support a greater variety of building blocks per coding position.

We have previously reported that PNA building blocks containing side chains polymerize efficiently and sequence specifically<sup>8</sup>. The addition of side-chain modified building blocks, especially those that display anionic or hydrophobic groups, could significantly increase the complexity and structural diversity of the resulting PNA library. PNA backbone modifications can induce a minimal to significant effect on DNA-PNA affinity<sup>8,11</sup>. This observation suggests that a prudent assignment of side-chain modifications to particular building blocks might increase the polymerization efficiency of poorly performing building blocks or decrease the polymerization efficiency of building blocks that are incorporated with unusual efficiency, allowing them to form an expanded set of building blocks without necessarily compromising the uniformity of DNA-templated polymerization. The incorporation of side chains on PNA backbones would also increase their functional potential, and the freedom to choose side-chain groups on the basis of their compatibility with templated polymerization rather than their ability to serve as substrates for biosynthetic machinery may enable the evolution of synthetic polymers containing functionality that has never been generated enzymatically.

In addition to the functional polymers that might arise from synthetic polymer evolution, research in this area will illuminate the relationship between building block structure, backbone structure, and the functional potential of a polymer. The transitions from a world based on a primitive polymer to a world based on RNA to a world based on proteins were mediated by information transfer steps that remain largely unknown. The translation of information-carrying polymers into a variety of other, more functional polymers could begin to generate plausible model systems for understanding these transitions. The continued development of key components towards the evolution of synthetic polymers may therefore give rise both to novel functional molecules, as well as to new insights into life's chemical beginnings.

## Methods

### PNA Building Block Preparation

PNA pentamer aldehydes were synthesized using preloaded H-Thr-Gly-NovaSyn TG resin (Novabiochem) and standard automated Fmoc solid-phase peptide synthesis on an Applied Biosystems 433A peptide synthesizer based on previously described protocols<sup>8,9</sup>. PNA pentamer aldehydes were purified by reverse-phase high-pressure liquid chromatography (HPLC).

### DNA-Templated PNA Polymerization

A typical polymerization reaction contained DNA template (0.4  $\mu\text{M}$ , when a known amount of template was used) and 16  $\mu\text{M}$  PNA aldehyde building blocks in 100 mM TAPS pH 8.5 buffer and 80 mM NaCl. Samples were heated briefly to 95 °C then cooled to reaction temperature. Reductive amination was initiated by the addition of freshly made 4 M aqueous sodium cyanoborohydride to a final concentration 80 mM. After incubation for 60 min, reactions were quenched by the addition of allyl amine to a final concentration of 260 mM. The reactions were purified by ethanol precipitation (Figs. 2–5) or with the Minelute Nucleotide Removal Kit (Qiagen) (Fig. 6) and, where appropriate, analyzed by denaturing polyacrylamide gel electrophoresis.

### PNA Strand Displacement

The product of PNA polymerization (to a final concentration of  $\sim 0.1 \mu\text{M}$ ) was combined with Herculase Buffer (Stratagene), and supplemented with dNTPs (to a final concentration of 2mM) and water to a final volume of 50  $\mu\text{L}$ . In a thermocycler, samples were denatured at 95 °C for 3 min and 1  $\mu\text{L}$  of Herculase Polymerase II (Stratagene) was added. Samples were incubated at 95 °C for 1 min to assure temperature equilibration in all samples following enzyme addition, then incubated for 3 min at 76 °C. Samples were purified using the Minelute Nucleotide Removal Kit (Qiagen). To assay PNA strand displacement and dsDNA formation by restriction digestion, 10  $\mu\text{L}$  of the unpurified material was combined with NEB Buffer #1 (to a final concentration of 1x), bovine serum albumin (to 100  $\mu\text{g}/\text{mL}$ ), and water to a volume of 96  $\mu\text{L}$ . *BccI* restriction endonuclease (4  $\mu\text{L}$ ) was added and the samples were incubated for 30 min at 37 °C. The digestion reactions were then purified using the Minelute Nucleotide Removal Kit (Qiagen).

## Streptavidin Binding Selections

The library of DNA templates or the positive control template (see Supplementary Methods for sequences) were separately amplified by PCR as described in the Supplementary Methods. The PCR products were quantified by UV spectroscopy and mixed together in a 1:100, 1:10,000, or 1:1,000,000 ratio favoring the library. The samples were subjected to strand separation (Supplementary Methods) and hairpin ligation (Supplementary Methods). PNA polymerization was performed in 50  $\mu$ L reactions in the presence 30  $\mu$ M total of the 12 PNA building blocks comprising the genetic code and 2  $\mu$ M of the biotinylated ggatt building block. A separate 1:100 ratio library also underwent selection, but lacked the biotinylated ggatt building block during the polymerization steps. The polymerization reactions were incubated for 1 hour at room temperature and purified using the Minelute Nucleotide Removal Kit (Qiagen). Samples were subjected to PNA strand displacement as described above.

Streptavidin-coated magnetic particles (50  $\mu$ L, Roche) were washed twice with 100  $\mu$ L of binding buffer (25mM Tris-HCl, 130mM NaCl, pH 7.4) and then incubated with 3 mg/mL yeast RNA (Ambion) in binding buffer for 30 min at room temperature. Strand displacement reaction samples were diluted 2-fold with binding buffer and added to the bead suspensions. Suspensions were incubated for 30 min with shaking (vortexing on the lowest setting using a Vortex Genie (Fisher)) at room temperature. The beads were washed twice with 100  $\mu$ L of binding buffer and once with 100  $\mu$ L of NEB Buffer #4. NEB Buffer #4 (100  $\mu$ L) was added to the beads and 1  $\mu$ L of *Ava*I restriction endonuclease (New England Biolabs) was added. Samples were incubated for 15 min at 37  $^{\circ}$ C. The beads were retained with a magnet and the solution (~100  $\mu$ L) was removed. From this solution, 4.5  $\mu$ L was diluted into 450  $\mu$ L of PCR amplification mix.

PCR amplification mix contained 450  $\mu$ L of iQ SYBR Supermix (Bio-rad), 1  $\mu$ M primer 1, 1  $\mu$ M primer 2 and water. Primer 1: 5' PCGAATTCCTGGCTCGGAAA where P is a 5' phosphate (Glen Research, 10-1901); Primer 2: 5' BGGCGTGCCCACTCGG where B is installed with the 5' Biotin phosphoramidite (Glen Research 10-5950). 50  $\mu$ L of the material was loaded onto a CFX-90 real-time PCR machine (Bio-Rad). An initial denaturation step at 94  $^{\circ}$ C for 3 min was followed by 40 cycles of [30 sec at 94  $^{\circ}$ C, 30 sec at 55  $^{\circ}$ C, and 30 sec at 72  $^{\circ}$ C]. A threshold cycle number ( $n$ ) was determined close to the inflection point of the qPCR fluorescence signal. 400  $\mu$ L of the material was amplified using a conventional PCR thermocycler. An initial denaturation step at 94  $^{\circ}$ C for 3 min was followed for  $n$  cycles of [30 sec at 94  $^{\circ}$ C, 30 sec at 55  $^{\circ}$ C, and 30 sec at 72  $^{\circ}$ C]. PCR reaction products were purified using the Minelute PCR purification kit (Qiagen) to remove primers, resulting in 10  $\mu$ L of purified PCR product per 400  $\mu$ L of PCR reaction. For further translation-selection-amplification cycles, this material was subjected to strand separation (Supplementary Methods). To distinguish the positive control template from the library templates, 8  $\mu$ L of the PCR reaction was combined with 0.5  $\mu$ L of *Msp*I restriction endonuclease and incubated for 30 min at 37  $^{\circ}$ C, then analyzed on a 2.5% agarose gel.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

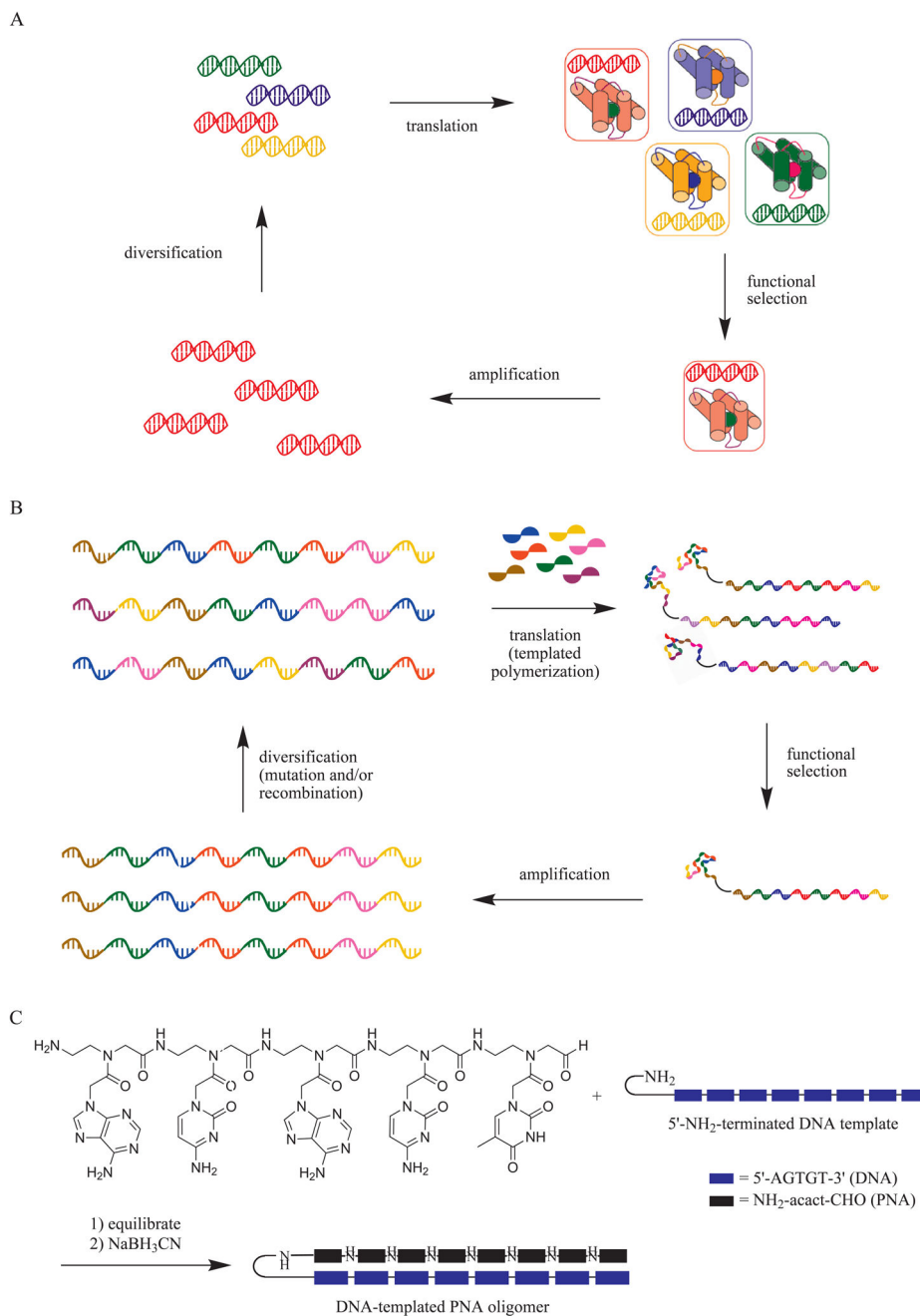
## Acknowledgments

This research was supported by the Office of Naval Research (N00014-03-1-0749), the NIH (R01GM065865), and the Howard Hughes Medical Institute. Y.B. gratefully acknowledges the support of an NSF Graduate Research Fellowship.

## References

1. Orgel LE. Unnatural selection in chemical systems. *Acc Chem Res.* 1995; 28:109–118. [PubMed: 11542502]
2. Leitzel JC, Lynn DG. Template-directed ligation: from DNA towards different versatile templates. *Chem Rec.* 2001; 1:53–62. [PubMed: 11893058]
3. Brudno Y, Liu DR. Recent progress toward the templated synthesis and directed evolution of sequence-defined synthetic polymers. *Chem Biol.* 2009; 16:265–76. [PubMed: 19318208]
4. Keefe AD, Cload ST. SELEX with modified nucleotides. *Curr Opin Chem Biol.* 2008; 12:448–456. [PubMed: 18644461]
5. Wang L, Xie J, Schultz PG. Expanding the genetic code. *Annu Rev Biophys Biomol Struct.* 2006; 35:225–49. [PubMed: 16689635]
6. Li X, Zhan ZY, Knipe R, Lynn DG. DNA-catalyzed polymerization. *J Am Chem Soc.* 2002; 124:746–7. [PubMed: 11817938]
7. Kozlov IA, De Bouvere B, Van Aerschot A, Herdewijn P, Orgel LE. Efficient transfer of information from hexitol nucleic acids to RNA during nonenzymatic oligomerization. *J Am Chem Soc.* 1999; 121:5856–9. [PubMed: 11542282]
8. Kleiner RE, Brudno Y, Birnbaum ME, Liu DR. DNA-templated polymerization of side-chain-functionalized peptide nucleic acid aldehydes. *J Am Chem Soc.* 2008; 130:4646–4659. [PubMed: 18341334]
9. Rosenbaum DM, Liu DR. Efficient and sequence-specific DNA-templated polymerization of peptide nucleic acid aldehydes. *J Am Chem Soc.* 2003; 125:13924–13925. [PubMed: 14611205]
10. Egholm M, et al. PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature.* 1993; 365:566–8. [PubMed: 7692304]
11. Englund EA, Appella DH. Gamma-substituted peptide nucleic acids constructed from L-lysine are a versatile scaffold for multifunctional display. *Angew Chem Int Ed Engl.* 2007; 119:1436–1440.
12. Datta B, Bier ME, Roy S, Armitage BA. Quadruplex formation by a guanine-rich PNA oligomer. *J Am Chem Soc.* 2005; 127:4199–4207. [PubMed: 15783201]
13. Sharma NK, Ganesh KN. PNA C-C+ i-motif: superior stability of PNA TC8 tetraplexes compared to DNA TC8 tetraplexes at low pH. *Chem Commun (Camb).* 2005:4330–2. [PubMed: 16113738]
14. Krishnan-Ghosh Y, Stephens E, Balasubramanian S. A PNA4 quadruplex. *J Am Chem Soc.* 2004; 126:5944–5. [PubMed: 15137741]
15. Krishnan-Ghosh Y, Stephens E, Balasubramanian S. PNA forms an i-motif. *Chem Commun (Camb).* 2005:5278–80. [PubMed: 16244727]
16. Cutrona G, et al. The peptide nucleic acid targeted to a regulatory sequence of the translocated c-myc oncogene in Burkitt's lymphoma lacks immunogenicity: follow-up characterization of PNAEmu-NLS. *Oligonucleotides.* 2007; 17:146–50. [PubMed: 17461771]
17. Nielsen PE. Addressing the challenges of cellular delivery and bioavailability of peptide nucleic acids (PNA). *Q Rev Biophys.* 2005; 38:345–50. [PubMed: 16600055]
18. Demidov VV, et al. Stability of peptide nucleic acids in human serum and cellular extracts. *Biochem Pharmacol.* 1994; 48:1310–3. [PubMed: 7945427]
19. Giesen U, et al. A formula for thermal stability (Tm) prediction of PNA/DNA duplexes. *Nucleic Acids Res.* 1998; 26:5004–5006. [PubMed: 9776766]
20. Griffin TJ, Smith LM. An approach to predicting the stabilities of peptide nucleic acid: DNA duplexes. *Anal Biochem.* 1998; 260:56–63. [PubMed: 9648653]

21. Sen A, Nielsen PE. Unique properties of purine/pyrimidine asymmetric PNA. DNA duplexes: differential stabilization of PNA.DNA duplexes by purines in the PNA strand. *Biophys J.* 2006; 90:1329–37. [PubMed: 16326919]
22. Ichida JK, et al. An *in vitro* selection system for TNA. *J Am Chem Soc.* 2005; 127:2802–2803. [PubMed: 15740086]
23. Cherny DY, et al. DNA unwinding upon strand-displacement binding of a thymine-substituted polyamide to double-stranded DNA. *Proc Natl Acad Sci USA.* 1993; 90:1667–70. [PubMed: 8383322]
24. Peffer NJ, et al. Strand-invasion of duplex DNA by peptide nucleic acid oligomers. *Proc Natl Acad Sci USA.* 1993; 90:10648–52. [PubMed: 8248156]
25. Smolina IV, Demidov VV, Soldatenkov VA, Chasovskikh SG, Frank-Kamenetskii MD. End invasion of peptide nucleic acids (PNAs) with mixed-base composition into linear DNA duplexes. *Nucleic Acids Res.* 2005; 33:e146. [PubMed: 16204449]
26. SantaLucia J Jr, Hicks D. The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct.* 2004; 33:415–40. [PubMed: 15139820]
27. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science.* 1990; 249:505–10. [PubMed: 2200121]
28. Doyon JB, Snyder TM, Liu DR. Highly sensitive *in vitro* selections for DNA-linked synthetic small molecules with protein binding affinity and specificity. *J Am Chem Soc.* 2003; 125:12372–3. [PubMed: 14531656]
29. Kanan MW, Rozenman MM, Sakurai K, Snyder TM, Liu DR. Reaction discovery enabled by DNA-templated synthesis and *in vitro* selection. *Nature.* 2004; 431:545–9. [PubMed: 15457254]
30. Wrenn SJ, Weisinger RM, Halpin DR, Harbury PB. Synthetic ligands discovered by *in vitro* selection. *J Am Chem Soc.* 2007; 129:13137–43. [PubMed: 17918937]
31. Scheuermann J, et al. DNA-encoded chemical libraries for the discovery of MMP-3 inhibitors. *Bioconjug Chem.* 2008; 19:778–85. [PubMed: 18254582]
32. Fitzwater T, Polisky B. A SELEX primer. *Meth Enzymol.* 1996; 267:275–301. [PubMed: 8743323]
33. Gold L, Polisky B, Uhlenbeck O, Yarus M. Diversity of oligonucleotide functions. *Annu Rev Biochem.* 1995; 64:763–97. [PubMed: 7574500]
34. Chen SJ. RNA folding: conformational statistics, folding kinetics, and ion electrostatics. *Annu Rev Biophys.* 2008; 37:197–214. [PubMed: 18573079]
35. Kuhn H, et al. Hybridization of DNA and PNA molecular beacons to single-stranded and double-stranded DNA targets. *J Am Chem Soc.* 2002; 124:1097–103. [PubMed: 11829619]
36. Seitz O. Solid-phase synthesis of doubly labeled peptide nucleic acids as probes for the real-time detection of hybridization. *Angew Chem Int Ed Engl.* 2000; 39:3249–3252. [PubMed: 11028066]
37. Davidson AR, Lumb KJ, Sauer RT. Cooperatively folded proteins in random sequence libraries. *Nat Struct Biol.* 1995; 2:856–64. [PubMed: 7552709]
38. Sabeti PC, Unrau PJ, Bartel DP. Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chem Biol.* 1997; 4:767–74. [PubMed: 9375255]
39. Joyce GF. Directed evolution of nucleic acid enzymes. *Annu Rev Biochem.* 2004; 73:791–836. [PubMed: 15189159]



**Figure 1. Strategies to evolve biological and synthetic polymers**

(a) Evolution of a biopolymer in nature or in the laboratory. Biological polymers are translated from and spatially associated with an information carrier (typically DNA). The biopolymers undergo selection based on their functional properties. The information encoding surviving biopolymers replicates and mutates inside or outside of cells, resulting in a second generation of biopolymer variants related to those that survived selection. (b) Synthetic polymers can in principle undergo a similar evolutionary process. Translation could be effected by enzymatic or non-enzymatic templated synthesis in a manner that associates each synthetic polymer with its information carrier. Following selection, the

information carriers encoding surviving synthetic polymers are amplified and mutated to generate templates for a subsequent round of translation. (c) Representative polymerization of a PNA aldehyde building block ( $\text{H}_2\text{N-acact-CHO}$ ) on a complementary DNA template. Fig. 1a and Fig. 1b are adapted from Ref. <sup>3</sup>.

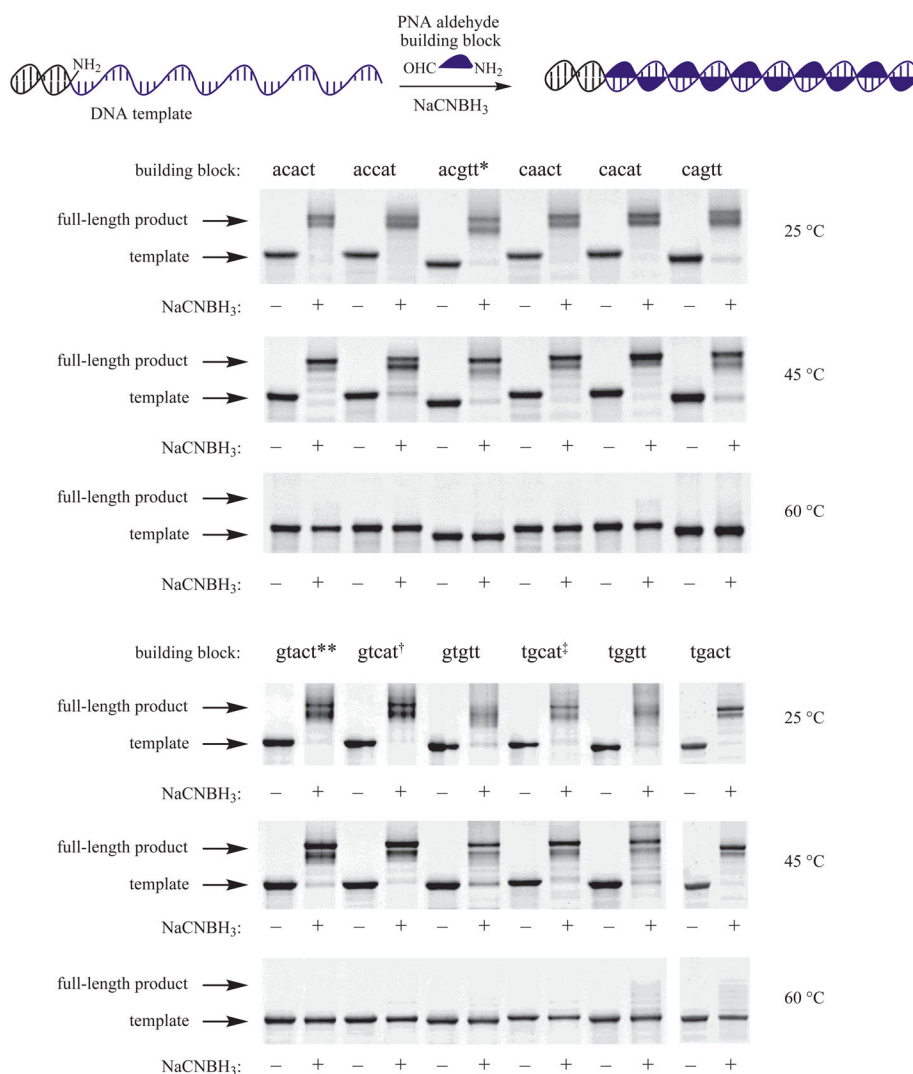
Author Manuscript

Author Manuscript

Author Manuscript

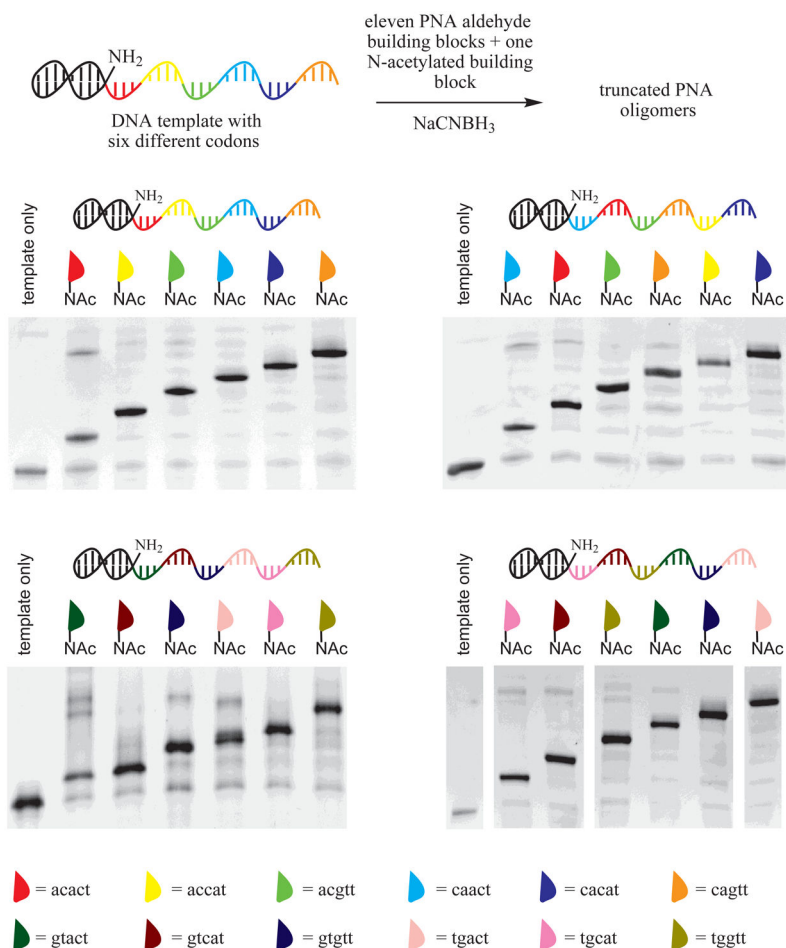
Author Manuscript





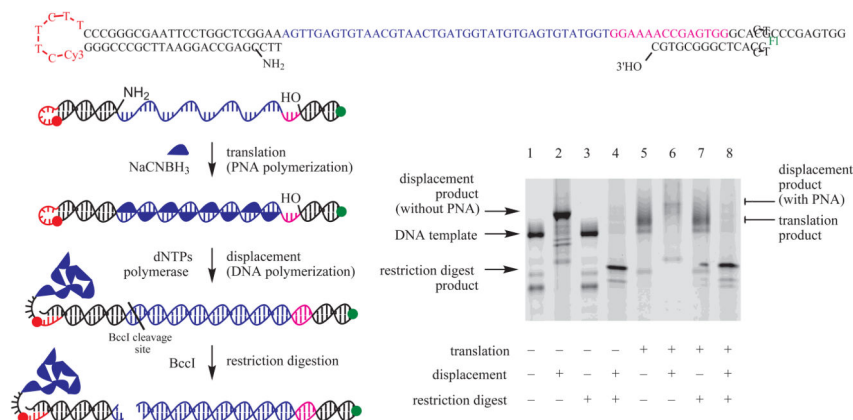
**Figure 2. A PNA genetic code that exhibits uniform DNA-templated polymerization**

Ten consecutive DNA-templated reductive amination couplings result in PNA 50-mers containing secondary amine linkages between every fifth nucleotide. The efficiency of this process is shown at 25 °C, 45 °C and 60 °C for the 12 PNA pentamer aldehydes shown. Reaction conditions: 16  $\mu\text{M}$  PNA aldehyde and 0.4  $\mu\text{M}$  DNA templates were heated to the specified temperature.  $\text{NaBH}_3\text{CN}$  was added to 80 mM. After 20 min, the reactions were analyzed by denaturing PAGE and visualized by staining with ethidium bromide. Except for the cases noted below in which repetitive template sequences adopted strong secondary structures, the coding region of the DNA templates contained ten consecutive repeats of the codon complementary to the building block shown. \*DNA template:  $(\text{AACGTAACAC})_5$ ; \*\*DNA template:  $(\text{AGTACAACCA})_5$ ; †DNA template:  $(\text{ATGCAAACCA})_5$ ; ‡DNA template:  $(\text{ATGCAAACCA})_5$ . In these four cases, 8  $\mu\text{M}$  of the assayed building block was supplemented with 8  $\mu\text{M}$  of the  $\text{H}_2\text{N-tggtt-CHO}$  building block.

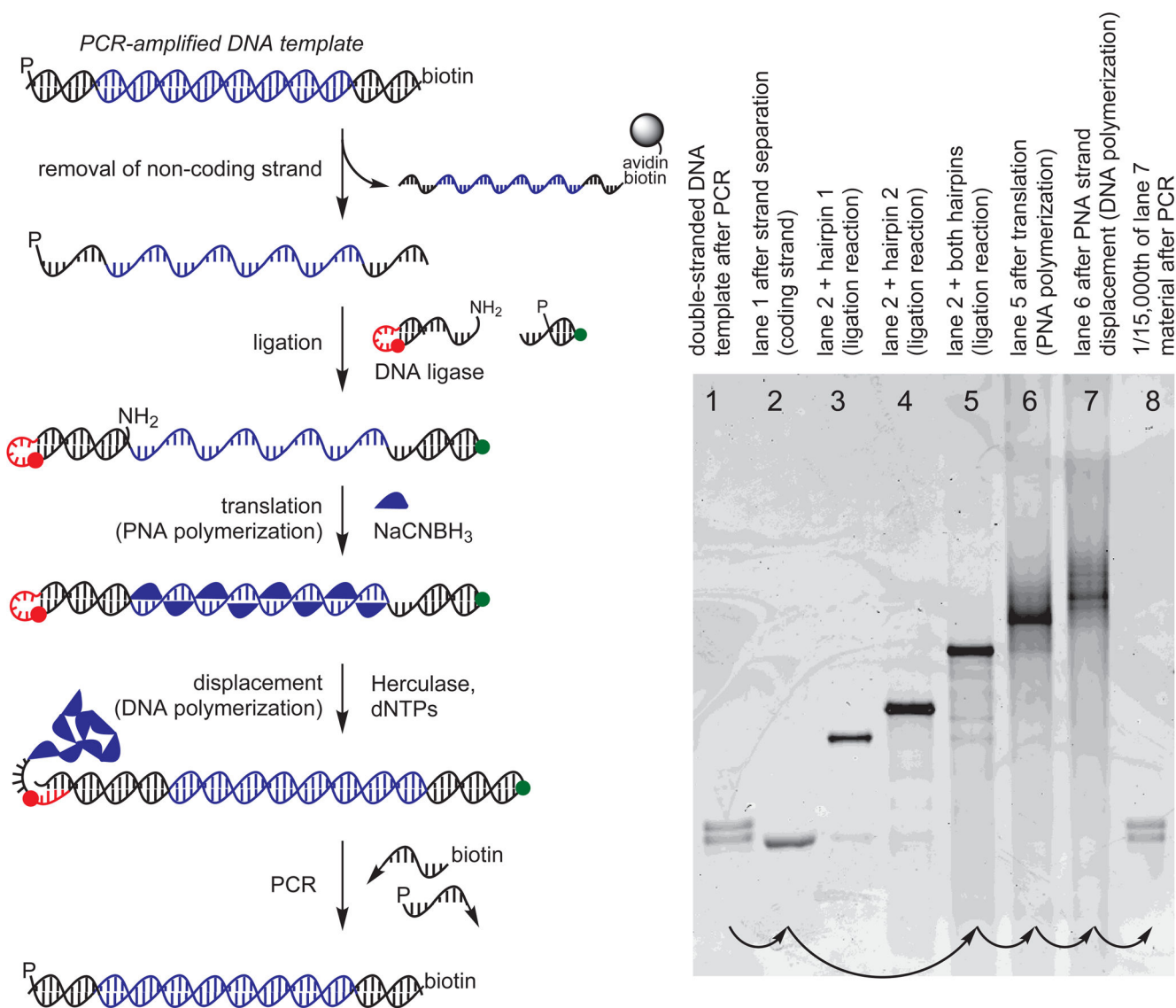


**Figure 3. Sequence-specific DNA-templated PNA polymerization using all 12 building blocks in a variety of contexts**

The four templates shown each contain six different codons and collectively use each of the 12 codons in two different contexts. Each gel lane shows a polymerization reaction containing 11 free amine PNA aldehyde building blocks and one capped (*N*-acetylated) PNA building block. Because the capped building block uniquely terminates polymerization, the length of the resulting PNA oligomer reveals the sequence specificity of the building block's coupling. Reaction conditions: 7  $\mu$ M of each PNA aldehyde and 0.4  $\mu$ M DNA templates were heated to 95  $^{\circ}$ C and cooled to 37  $^{\circ}$ C. NaBH<sub>3</sub>CN was added to 80 mM. After 60 min at 37  $^{\circ}$ C the reactions were quenched and analyzed by denaturing PAGE and visualized by staining with ethidium bromide.



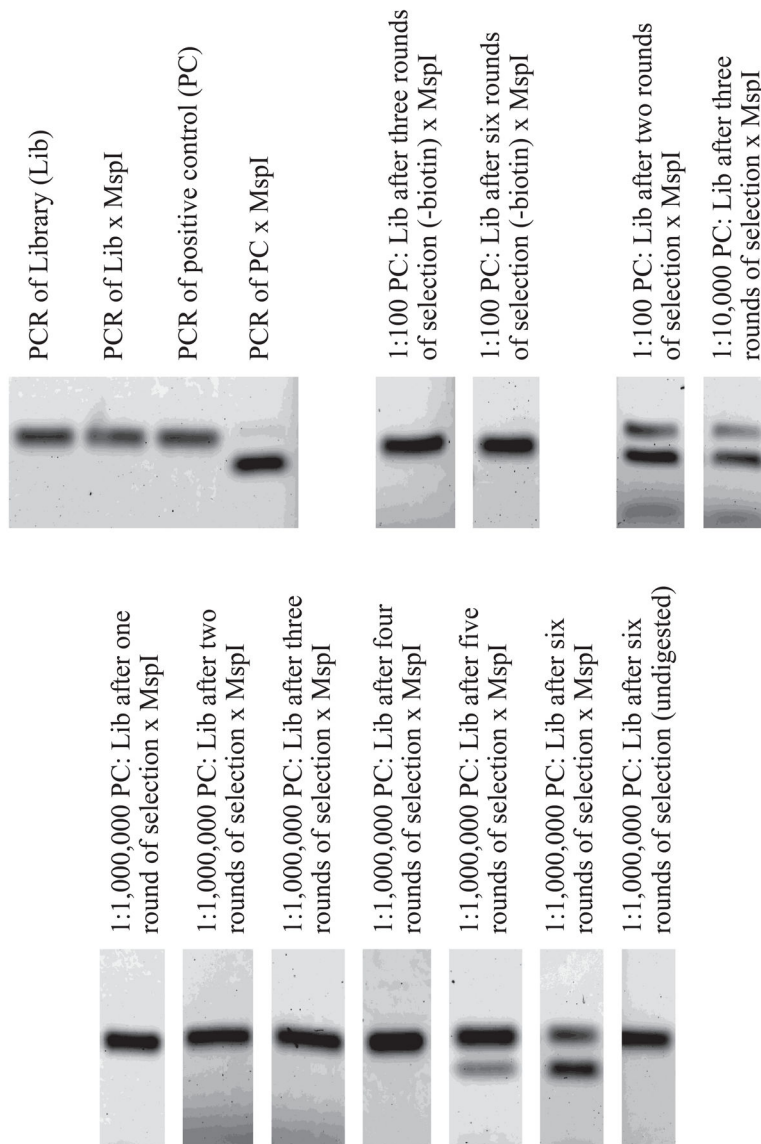
**Figure 4. Translation of a DNA template into PNA and displacement of the resulting PNA strand** Following PNA polymerization (lane 5) under conditions described in Figure 2, the PNA strand is displaced by primer extension of the 3' hairpin using Herculase DNA polymerase II and dNTPs (lane 6). DNA polymerization to the end of the template strand (through the 5' stem-loop) generates an 84-base pair DNA duplex that can form as a mutually exclusive alternative to the 40-base pair PNA-DNA heteroduplex and 24-base pair DNA-DNA stem. Successful strand displacement creates a double-stranded *BccI* restriction endonuclease cleavage site. The *BccI*-mediated cleavage of the primer extension product (lanes 4 and 8) is therefore consistent with displacement of the PNA strand. The red dot represents a Cy3 backbone spacer that blocks subsequent DNA polymerization. The green dot represents a fluorescein group. The denaturing PAGE gel shown is visualized by imaging fluorescein fluorescence. Reactions shown in lanes labeled “translation–” lacked NaCNBH<sub>3</sub>; those in lanes labeled “displacement–” lacked DNA polymerase; those in lanes labeled “restriction digestion–” lacked *BccI*.



**Figure 5. A full cycle of translation, displacement, simulated selection, and PCR amplification for a single DNA template encoding a PNA 40-mer**

Double-stranded DNA template (lane 1) was immobilized on streptavidin-linked beads and the non-biotinylated coding strand (lane 2) was eluted. Two hairpins were ligated onto the coding strand (lane 5) to generate a template ready for translation into PNA (lane 6) and displacement of the resulting PNA strand (lane 7). A small amount (1/15,000th) of the translated and displaced material was subjected to PCR, regenerating the double-stranded template (lane 8). Lanes 3 and 4 show hairpin ligation reactions with only one hairpin. The denaturing PAGE gel shown was visualized by ethidium bromide staining.

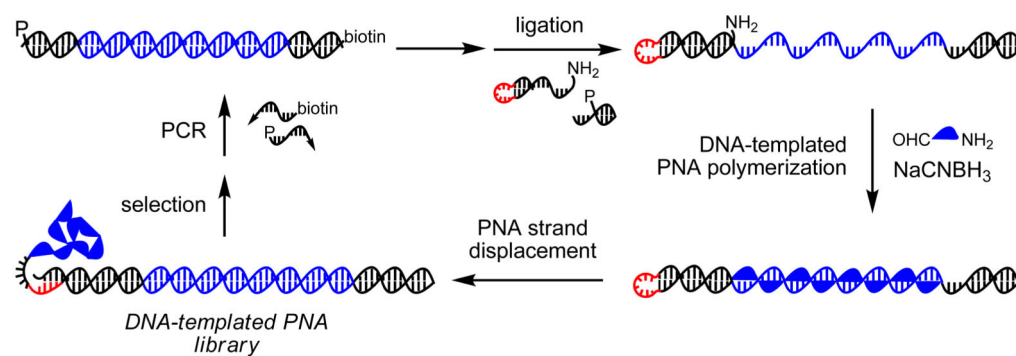
library (Lib)  
 coding region: 5'-AY'X'AY'X'AY'X'AY'X'AY'X'AY'X'AY'X'AY'X'AA-3'  
 positive control (PC)  
 coding region: 5'-AGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAATCCGG-3'  
MspI cleavage site



**Figure 6. Model selection of a DNA-templated PNA 40-mer library**

To a library of DNA templates encoding  $4.3 \times 10^8$  PNA 40-mers containing eight consecutive building blocks was added  $10^{-2}$ ,  $10^{-4}$ , or  $10^{-6}$  equivalents of a DNA template uniquely containing an AATCC codon (red) that translates into a biotinylated H<sub>2</sub>N-ggatt-CHO PNA building block. The positive control DNA template, but not the other templates, contains an *MspI* cleavage site. The resulting library of DNA templates was subjected to multiple iterated rounds of translation, PNA strand displacement, *in vitro* selection for binding to immobilized streptavidin, and amplification by PCR (Fig. 5). *MspI* digestion of

the PCR-amplified DNA before and after selection reveals that the selection resulted in > 1,000,000-fold enrichment for the positive control sequence encoding the biotinylated PNA. A similar model selection using  $10^{-2}$  equivalents of the positive control template but omitting the biotinylated building block during translation (-biotin lanes) results in no enrichment of the positive control sequence.  $Y'$  = mixture of AC, GT, and TG;  $X'$  = mixture of AC, GT, TG, and CA. The images shown are from 2.5% agarose gels stained with ethidium bromide. The complete set of gel images analyzing all model selections are provided in Supplementary Fig. 3.



**Figure 7.**