

Using Synthetic Nucleic Acids for *in vivo* System Assembly

Austin Che

December 16, 2003

ABSTRACT: Many types of synthetic biological molecules have been designed that are similar to nucleic acids. We propose engineering variants of the replication enzymes so that a cell is able to copy and maintain a synthetic nucleic acid *in vivo*. In addition, the cell should not be able to transcribe the synthetic nucleic acid into RNA. Developing these new enzymes would allow the replication but not the transcription of constructs located on plasmids containing the nucleic acid analog. Thus, common and familiar molecular biology techniques can be maintained even while propagating constructs that are toxic to the cell.

1 Motivation

1.1 System Assembly

For *synthetic biology* [26] to succeed, it is necessary to assemble arbitrarily complex biological systems. To assemble systems, it is necessary to be able to stably propagate partial assemblies on plasmids. Until we can easily synthesize arbitrarily long sequences of DNA, propagation and assembly will continue to occur in cells, notably in *E. coli*. Even DNA synthesis companies, such as Blue Heron, that claim to be able to synthesize long pieces of DNA, propagate the synthesized DNA on plasmids in *E. coli*. Our experience has shown that Blue Heron has problems synthesizing strong promoters or sequences designed to be unfavorable for *E. coli*. These sequences show a strong tendency to mutate and, therefore, cannot be stably maintained. There should not be unnecessary pressures for cells to mutate a system we are trying to build. Thus, during assembly, the system should have minimal effect on the cell. If possible, we would like any insert carried on a plasmid to not be expressed inside the cell.

Biologists have for decades been building simple constructs and propagating them in cells and have had to deal with constructs that are toxic to the

host. The general approach has been to use low-copy plasmids or an inducible promoter, e.g. P_{lac} , that only turns on expression at a desired point, allowing for the stable maintenance of the plasmid and insert. However, the inducible promoter approach is neither general nor sufficient for any slightly more complex system, which may contain many components and promoters that cannot be regulated easily.

Another approach may be to use an *in vitro* replication system [8]. But *in vitro* replication eliminates many of the benefits and the built-up familiarity with cells. Some may propose assembling in a different organism from the final target organism, e.g. maintaining plasmids containing *E. coli* parts in yeast. But, regardless of the organism used, at some point, one will want to assemble components without interference with the native cell. In addition, nothing is as well-understood and as easily manipulated as *E. coli*. A more satisfying solution would allow for the assembly of any system in any organism.

1.2 Proposal Overview

We need a cloning vector, such as a plasmid, that would be copied by the cell during division but whose insert would not be expressed. One way would be for the vector to be recognized by DNA polymerase

but not RNA polymerase, ensuring faithful copying during replication and limited effects on the cell due to no transcriptional expression of the system.

To achieve the previous goals, we would like to be able to use a *synthetic nucleic acid* (SNA) *in vivo*. SNA will be used here to represent one member from the entire family of nucleic acid analogs. The desired properties for the SNA include:

- The SNA must be able to be easily replicated and maintained in cells. This means that our cloning strain must contain a DNA polymerase-like enzyme that can faithfully replicate SNA to produce more SNA.
- The SNA must not be recognized by RNA polymerase, as SNA inserts should not be expressed.
- The SNA should be easily convertible to and from DNA to allow compatibility with other techniques.
- Ideally, a plasmid can consist of a mix of DNA and SNA. We would like to maintain a positive selection for the plasmid, and, for this, maintaining expression of an antibiotic resistance gene would be helpful.

An appropriate SNA satisfying the above properties would allow for propagation of assemblies as SNA. The cells would replicate but not express the system. When expression is desired, the SNA can be converted to DNA. To achieve the above objectives, it is necessary to select the form of the SNA and design the appropriate enzymes.

2 Proposed Methodology

2.1 Overview

The desired goals can be broken down into several steps:

1. Choose a particular SNA.
2. Try to engineer a DNA polymerase variant that can replicate the SNA *in vitro*.

3. See if the DNA polymerase can be made to work *in vivo*.
4. Check that RNA polymerase does not recognize the SNA.
5. Try a different SNA if the above does not work.

2.2 Synthetic Nucleic Acids

There have been many artificially designed chemical molecules that resemble DNA, each with its own properties [7]. Changes to nucleic acids can be made in the bases used (other than A, T, U, C, G), the sugar (other than ribose or deoxyribose), or the backbone (other than a phosphodiester bond). The difficult part is not coming up with a nucleic acid analog, but rather coming up with one that has a reasonable chance of working *in vivo*.

2.2.1 Nitrogenous Bases

It may be possible to introduce new complementary bases, in addition to the 4 natural bases. Ideally, the synthetic bases would not be able to base pair with the natural bases and could only pair with the other synthetic bases. Recently, all four base pairs were replaced with larger pairs that formed a more stable and wider helix [16]. However, these could still base pair with the native bases, making it not too useful for this application.

Using other bases would be the easiest approach if it were possible to come up with 4 new, suitable bases. If the synthetic base pairs take up the same amount of space as the normal base pair, then the DNA polymerase should function correctly as long as the required dNTPs are available, with the benefit that it would be relatively straightforward for the DNA polymerase to copy both the natural and the synthetic bases. Preventing RNA polymerase from transcribing the synthetic base pairs is also straightforward. If the cell is not allowed to make any of the synthetic rNTP version of the synthetic bases, then transcription should be stopped.

Twelve bases that can form six mutually exclusive base pairs have been described and could in principle be used for our SNA [25]. In addition, some of these

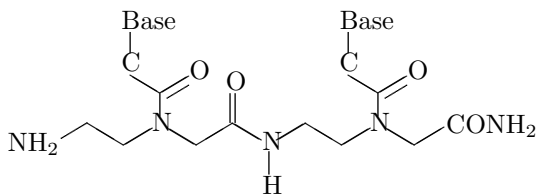


Figure 1: Peptide Nucleic Acid (PNA) consists of a peptide like backbone joining the purines and pyrimidines of DNA, allowing it to base pair with DNA.

bases have been shown to be extended, albeit with lower efficiency, by some DNA polymerases, with the Klenow fragment working the best. One pair of artificial bases has been shown to be incorporated by DNA polymerase with high fidelity [22]. Other synthetic bases have been used to expand the genetic code and to incorporate non-natural amino acids during translation [2]. These results show that, for some bases, many enzymes may just work without needing extra manipulations.

2.2.2 PNA

Peptide Nucleic Acid (PNA) was designed by replacing the phosphate backbone of DNA with N-(2-amino-ethyl)glycine and connecting the base by a methylenecarbonyl group [19]. PNA has a peptide like backbone but uses the bases found in DNA. It was designed to fit in roughly the same space as DNA, allowing PNA to form strong base pairing with DNA. PNA is relatively easy to synthesize with manual methods or on standard DNA or peptide synthesizers [6]. Figure 1 shows the structure of PNA.

PNA shows strongest binding with PNA itself but can also bind with complementary DNA or RNA according to the standard Watson-Crick rules [27]. The PNA backbone is uncharged, making the PNA:DNA duplexes more stable than DNA:DNA due to a lack of electrostatic repulsion. PNA oligos can block the transcription machinery but not the DNA replication process [9]. Replication is thus more difficult to block than transcription, which is promising for our proposed application. PNA is highly stable in bio-

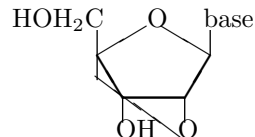


Figure 2: Locked Nucleic Acid contains an extra linkage between the 2'-O and 4'-C but is otherwise identical to RNA.

logical systems, showing resistance to nucleases and proteases [27].

PNA by itself is not recognized by nucleic acid enzymes such as DNA and RNA polymerases [27]. However, a PNA molecule with one normal nucleotide at the end has been shown to work as a primer for DNA polymerase [17]. Some DNA polymerases can elongate PNA primers much better than others. The best polymerases include the Klenow fragment from *E. coli* and the *Tth* polymerase.

2.2.3 LNA

Locked Nucleic Acid (LNA) consists of an extra methylene linkage between the 2'-oxygen and the 4' carbon of the ribose sugar [5] (Figure 2). This extra conformation restriction increases the binding of LNA with complementary DNA. LNA is more similar to DNA than PNA is. LNA has the same charged phosphodiester backbone as standard DNA, making it more soluble than PNA. LNA oligos can be also synthesized using the standard phosphoramidite chemistry and it is possible to synthesize chimeric LNA/DNA oligos [21].

2.2.4 Other Possibilities

Some other possibilities for nucleic acid analogs include: using left-handed DNA, methylated DNA, or DNA that uses L-ribose or perhaps a completely different sugar. For example, hexitol nucleic acids (HNA) [15] have a six-membered sugar ring and can form similar helical structures with normal nucleic

acids. Other nucleic acids analogs with different sugars have been shown to form much stronger base pairs than RNA and have been postulated to be part of the early evolution of nucleic acids [4, 10]. Thus, there is hope that a non-ribose-based nucleic acid could function and may even function better than the existing system.

2.3 DNA Polymerase

Although it is not an easy task to make a polymerase that can faithfully replicate a foreign type of nucleic acid, polymerases have been studied in depth [24] and we have good crystal structures of several DNA polymerases [11].

With the crucial role of DNA polymerases, it is somewhat surprising the diversity of polymerase structures [24]. This indicates that there are multiple ways of catalyzing the same reaction, giving hope that we can engineer a new DNA polymerase. All the polymerases share a similar structure with regions termed the palm, fingers, and thumb domains. The palm domains are relatively conserved across polymerase and helps catalyze the phosphoryl transfer reaction. The other two domains show less conservation.

Interactions such as hydrogen bonds between the polymerase and the substrate DNA and dNTPs are roughly known [11]. Binding of the correct dNTP rotates the fingers domain of the polymerase, which is necessary for catalysis. Three residues of the polymerase finger Arg 482, Lys 486, and Lys 560 catalyze the transfer of the dNTP to the primer. The incorporation of the correct dNTP depends only on the correct hydrogen pairing with the template strand. Thus, the polymerase is not affected too much by the actual base used, but rather only the location of the paired bases. The DNA polymerase makes many contacts with the phosphodiester backbone, but many of the residues involved are not conserved across polymerases. The backbone interactions thus do not seem critical for the polymerization activity.

As expected, DNA polymerases normally copy DNA with high fidelity. Some principles underlying this fidelity are known [3]. One of the goals is to lower the fidelity of the DNA polymerase for nat-

ural dNTPs so that it can incorporate other types of substrates. However, ideally, the engineered DNA polymerase should also have high fidelity for the SNA and be selective for SNA over DNA. It would be disastrous if the cell's chromosome was replicated into SNA, inhibiting transcription and gene expression.

2.4 Implementation

For choosing a particular SNA, LNA seems to be a good place to start. LNA is quite similar to DNA, with only a modification to the sugar portion that does not affect the chemistry of the polymerization reaction. The normal phosphodiester backbone in LNA makes it unnecessary to change the catalytic reaction but only the binding specificity.

For PNA, it would be necessary to make a DNA polymerase create peptide bonds. Of course, we have high resolution structures of ribosomes, but it is probably easier to make a DNA polymerase catalyze a peptide reaction than to make a ribosome/tRNA that can incorporate nucleic acid bases. Another place in nature from which to draw inspiration is from non-ribosomal peptide synthetases that can synthesize peptide-like molecules in cells [23]. This class of enzymes may provide insights into designing something that can catalyze the polymerization of something like PNA without the entire complexity of ribosomes.

Whatever the SNA is, not only does the polymerase need to be able to replicate it, the cell also either needs to be able to make the monomers needed for incorporation or the monomers can be provided in the medium if the cell can uptake the monomers. A side benefit of providing the SNA monomers in the medium is that the cell would no longer be able to independently replicate outside of the artificial system. Thus, there would be little chance of harm if it is accidentally released into the environment.

The process of redesigning the enzyme will probably be a mix of rational design, *in vitro* selection, and protein evolution to find what works. Re-designing polymerases with novel catalytic activities has been done before. There have been DNA polymerases rationally engineered to incorporate other nucleic acids such as ribonucleotides [12]. A directed evolution approach using phage display has been used to evolve

DNA polymerase into an extremely efficient RNA polymerase [28]. Another example is the commercially available Terminator DNA polymerase (New England Biolabs), which was engineered from a native DNA polymerase to more efficiently incorporate acyclonucleotides, a synthetic dNTP analog [13]. However, not only do we need a DNA polymerase that has increased specificity for our SNA, but also decreased specificity to incorporate normal DNA.

Many of the components that determine how the polymerase chooses the correct sugar have been studied [14]. We can learn from previous mutational studies of DNA polymerases. For example, to allow DNA polymerase to incorporate ribonucleotides requires changing a single residue that sterically blocks the extra 2'-OH of the ribose [1, 20]. As LNA contains the 2'-OH of ribose, we would expect that the same mutation would also be helpful for allowing LNA to enter the binding site.

The DNA polymerase may not be the only enzyme that needs to be engineered as there are many other enzymes involved in replication. If we use something like LNA, then there is a chance that DNA ligase and many of the other enzymes would still work. But it may be necessary to make a primase work with the synthetic nucleic acid to incorporate an appropriate primer for replication to succeed.

As an additional feature, it would be useful to have SNA/DNA hybrid molecules. The plasmid backbone could consist of DNA with the insert as SNA. We do not want to inhibit the expression of the entire plasmid and would like to continue to express an antibiotic at least. To do this would just require the ability to ligate between DNA and SNA during replication. For something like LNA, this is not much of an issue. For PNA, while PNA/DNA hybrids are easily chemically synthesized, it is unclear how to make an enzyme make a hybrid *in vivo*.

Finally, for usability and compatibility with other biology techniques, it is necessary to have a method to convert SNA to DNA and a way to convert DNA to SNA. This conversion could be done *in vitro* and may require engineering a polymerase that recognizes both SNA and DNA but is willing to incorporate any nucleotides that are in the environment. For something like LNA, it may be as simple as relaxing

the polymerase specificity so that it recognizes either LNA or DNA. By providing LNA as the template and dNTPs, then the conversion to normal DNA can be done easily. It may be difficult to do this conversion to normal DNA if non-natural bases are used, as it would require a type of translation that is more complicated than the standard polymerase activity. This is one potential reason to use standard bases in our SNA, even if it requires modifying some enzymes.

3 Conclusion

Assembling complex biological systems consisting of parts that may interact adversely with the host cell is currently not an easy task. The goal is to select a biologically-inspired nucleic acid analog that can be replicated by DNA polymerase but not transcribed by RNA polymerase, allowing for assembly and propagation with minimal effect on the host cell. It is unlikely that there exists a synthetic nucleic acid that would satisfy these requirements using native polymerases. However, we have the ability to engineer new enzymes through rational design and evolutionary approaches. Although it is undoubtedly a non-trivial task to generate an enzyme that can faithfully replicate a synthetic nucleic acid *in vivo*, the benefits to be gained are also non-trivial.

References

- [1] Mekbib Astatke, Kimmie Ng, Nigel D. F. Grindley, and Catherine M. Joyce. A single side chain prevents *Escherichia coli* DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc. Natl. Acad. Sci. USA*, 95(7):3402–3407, 1998.
- [2] J. D. Bain, Christopher Switzer, A. Richard Chamberlin, and Steven A. Benner. Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. *Nature*, 356:537–539, April 1992.

- [3] William A. Beard and Samuel H. Wilson. Structural insights into the origins of DNA polymerase fidelity. *Structure*, 11:489–496, May 2003.
- [4] M. Beier, F. Reck, T. Wagner, R. Krishnamurthy, and A. Eschenmoser. Chemical etiology of nucleic acid structure: Comparing pentopyranosyl-(2'→4') oligonucleotides with RNA. *Science*, 283:699–703, January 1999.
- [5] Dwaine A. Braasch and David R. Corey. Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. *Chemistry & Biology*, 8:1–7, 2001.
- [6] Dwaine A. Braasch and David R. Corey. Synthesis, analysis, purification, and intracellular delivery of peptide nucleic acids. *Methods*, 23:97–107, 2001.
- [7] Dwaine A. Braasch and David R. Corey. Novel antisense and peptide nucleic acid strategies for controlling gene expression. *Biochemistry*, 41(14):4503–4510, apr 2002.
- [8] Elliott Crooke. DNA synthesis initiated at oriC: *in vitro* replication reactions. In Judith L. Campbell, editor, *Methods in Enzymology*, volume 262, pages 500–506. Academic Press, 1995.
- [9] Silvia Diviacco, Valentina Rapozzi, Luigi Xodo, Claude H el ene, Franco Quadrifoglio, and Carine Giovannangeli. Site-directed inhibition of DNA replication by triple helix formation. *FASEB Journal*, 15:2660–2668, December 2001.
- [10] Albert Eschenmoser. Chemical etiology of nucleic acid structure. *Science*, 284(5423):2118–2124, June 1999.
- [11] Matthew C. Franklin, Jimin Wang, and Thomas A. Steitz. Structure of the replicating complex of a pol α family DNA polymerase. *Cell*, 105(5):657–667, June 2001.
- [12] Andrew F. Gardner and William E. Jack. Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Research*, 27(12):2545–2553, 1999.
- [13] Andrew F. Gardner and William E. Jack. Acyclic and dideoxy terminator preferences denote divergent sugar recognition by archaeon and *Taq* DNA polymerases. *Nucleic Acids Research*, 30(2):605–613, 2002.
- [14] Catherine M. Joyce. Choosing the right sugar: How polymerases select a nucleotide substrate. *Proc. Natl. Acad. Sci. USA*, 94:1619–1622, 1997.
- [15] E Lescrier, R Esnouf, J Schraml, R Busson, HA Heus, CW Hilbers, and P Herdewijn. Solution structure of a HNA-RNA hybrid. *Chemistry & Biology*, 7(9):719–731, September 2000.
- [16] Haibo Liu, Jianmin Gao, Stephen R. Lynch, Y. David Saito, Lystranne Maynard, and Eric T. Kool. A four-base paired genetic helix with expanded size. *Science*, 302(5646):868–871, October 2003.
- [17] M.J. Lutz, D.W. Will, G. Breipohl S.A. Benner, and E. Uhlmann. Synthesis of a monocharged peptide nucleic acid (PNA) analog and its recognition as substrate by DNA polymerases. *Nucleosides Nucleotides*, 18(3):393–401, 1999.
- [18] Peter E. Nielsen. Peptide nucleic acid: a versatile tool in genetic diagnostics and molecular biology. *Current Opinion in Biotechnology*, 12:16–20, 2001.
- [19] Peter E. Nielsen, Michael Egholm, Rolf H. Berg, and Ole Buchardt. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science*, 254(5037):1497–1500, December 1991.
- [20] Premal H. Patel and Lawrence A. Loeb. Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *Journal of*

- Biological Chemistry*, 275(51):40266–40272,
December 2000.
- [21] Michael Petersen and Jesper Wengel. LNA: a versatile tool for therapeutics and genomics. *TRENDS in Biotechnology*, 21(2):74–81, February 2003.
- [22] Joseph A. Piccirilli, Tilman Krauch, Simon E. Moroney, and Steven A. Benner. Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature*, 343:33–37, January 1990.
- [23] Stephan A. Sieber and Mohamed A. Marahiel. Learning from nature’s drug factories: Nonribosomal synthesis of macrocyclic peptides. *Journal of Bacteriology*, 185(24):7036–7043, December 2003.
- [24] Thomas A. Steitz. DNA polymerases: Structural diversity and common mechanisms. *Journal of Biological Chemistry*, 274(25):17395–17398, June 1999.
- [25] Christopher Y. Switzer, Simon E. Moroney, and Steven A. Benner. Enzymic recognition of the base pair between isocytidine and isoguanosine. *Biochemistry*, 32(39):10489–10496, 1993.
- [26] Synthetic Biology.
<http://www.syntheticbiology.org/>.
- [27] Eugen Uhlmann. Peptide nucleic acids (PNA) and PNA-DNA chimeras: from high binding affinity towards biological function. *Biological Chemistry*, 379:1045–1052, 1998.
- [28] Gang Xia, Liangjing Chen, Takashi Sera, Ming Fa, Peter G. Schultz, and Floyd E. Romesberg. Directed evolution of novel polymerase activities: Mutation of a DNA polymerase into an efficient RNA polymerase. *Proc. Natl. Acad. Sci. USA*, 99(10):6597–6602, May 2002.