

# Single tube gene synthesis by phosphoramidate chemical ligation

Afaf H. El-Sagheer,<sup>a,b</sup> and Tom Brown.<sup>a\*</sup>

Received 00th January 20xx,  
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

**Templated chemical ligation of 5'-amino and 3'-phosphate oligonucleotides was used to synthesise a 762 base pair gene for green fluorescent protein. The phosphoramidate linkage can be read by DNA polymerase and transcribed to make RNA. We also show that phosphoramidate ligation and orthogonal CuAAC-mediated DNA ligation can be used simultaneously.**

Gene synthesis is an important and rapidly growing field. The most successful and by far most common methods are based on PCR amplification of synthetic oligonucleotide pools. This methodology is used routinely to produce large DNA constructs up to several kilobases (kb) in length, and has served the biological community well.<sup>1</sup> However, it has limitations, foremost of which is its inability to produce DNA that contains modifications at specific pre-defined loci. Such modified DNA constructs, if available, would be useful in many applications including epigenetics.<sup>2</sup> The limitations of PCR are due to the fact that DNA polymerases cannot discriminate between the canonical deoxyribonucleoside triphosphates (dNTPs) and modified versions. This is because any modified dNTP must possess the same fundamental Watson-Crick base pairing properties as its natural counterpart in order to be incorporated into DNA by polymerase enzymes. Consequently the natural and unnatural dNTPs compete in an uncontrollable manner. An obvious solution to this problem is to assemble DNA by ligation of pre-synthesized chemically modified oligonucleotides. This would open up new areas of biology, allowing a vast array of modifications to be incorporated into genomic DNA.

Ligation can be carried out enzymatically,<sup>3</sup> but chemical

ligation offers an attractive alternative.<sup>4</sup> It is compatible with large scale applications, radical modifications to the sugars and nucleobases, templated or non-templated reactions, and can be carried out in conditions under which ligase enzymes would not remain functional, including automated nucleic acid assembly. Moreover, chemical ligation is not restricted to the natural phosphodiester backbone of DNA; other backbones can be produced, some of which are remarkably biocompatible.<sup>5,6</sup> The use of chemical ligation to produce modified backbones has other advantages; highly efficient chemical reactions can be chosen, orthogonal ligation chemistries can be used simultaneously for special applications (Figure 1A), and successful DNA ligation strategies, once developed, can be applied to the synthesis of long RNA strands<sup>7</sup> with potential applications in gene editing.<sup>8</sup>

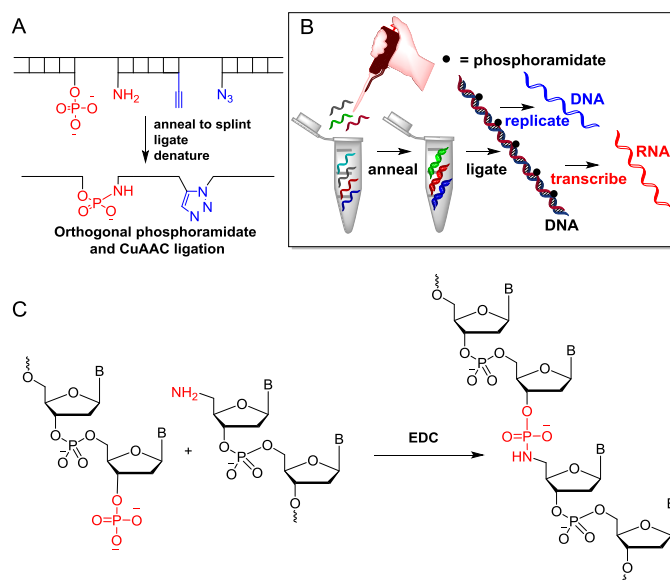


Figure 1. A) Templated simultaneous orthogonal phosphoramidate and CuAAC ligation reactions. B) The concept of single tube gene assembly by phosphoramidate ligation followed by transcription of modified DNA. C) Details of 5'-Phosphoramidate ligation chemistry.

We have previously used click chemistry to synthesise long DNA strands, and have shown that the triazole linkage at the site of ligation is compatible with DNA and RNA polymerase

<sup>a</sup> Department of Chemistry, University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford, OX1 3TA, UK.

<sup>b</sup> Chemistry Branch, Dept. of Science and Mathematics, Faculty of Petroleum and Mining Engineering, Suez University, Suez, 43721, Egypt.

\*author for correspondence

† Footnotes relating to the title and/or authors should appear here.

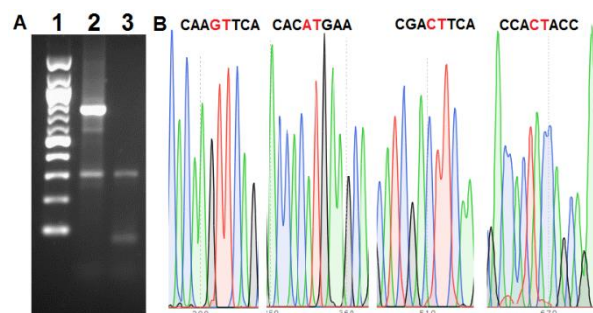
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

enzymes.<sup>7,9-11</sup> However, it would also be useful to have an orthogonal DNA ligation chemistry. In this regard, the 5'-phosphoramidate linkage<sup>12,13</sup> is particularly attractive. It can be made by ligation of oligonucleotides bearing simple 3'-phosphate and 5'-amino modifications and the ligation reaction can be catalysed by coupling agents that are commonly used in peptide synthesis (Figure 1b,c). Importantly, for any modified DNA ligation chemistry to be useful in biology, it should be compatible with the synthesis of functional genes, i.e. as well as being efficient, it must not give rise to mutations, and the modified linkage should be compatible with polymerase enzymes. This has not yet been proven for phosphoramidate ligation, although it has recently been used in an ingenious nanocage approach to the synthesis of DNA nanoconstructs.<sup>14</sup> Here we show, by constructing the gene for green fluorescent protein (EGFP), that DNA synthesis by phosphoramidate ligation is an efficient process that does not lead to corruption of the genetic information.

To evaluate phosphoramidate ligation in the context of gene assembly we initially synthesised an 81-mer DNA strand (ODN 1) with a phosphoramidate linkage at the ligation point by reacting 23-mer ODN 2 bearing a 3'-phosphate with a 5'-amino modified 58-mer ODN 3 (oligonucleotide sequences in SI Table 2). The reaction was templated with a complementary 18-mer DNA splint (ODN 4), and the individual strands were annealed by heating to 95 deg. C, then cooling slowly to room temperature before addition of the coupling agent. Reactions were carried out in HEPES buffer at pH 7.2 and 0.4M sodium chloride in the presence of EDC and 1-(2-hydroxyethyl) imidazole (SI Figure S1). Under these conditions the reaction was complete within 45 min (SI Figure S2). A PCR product was generated from the synthetic 81-mer using GoTaq DNA polymerase to show that the enzyme can efficiently read through the phosphoramidate linkage. Agarose gel-electrophoresis (SI Figure S3) and Sanger sequencing were then carried out on the PCR product to determine that it had the correct size and nucleobase sequence. Having established suitable ligation conditions, a longer DNA strand with two phosphoramidate linkages was constructed (303-mer, ODN 5) from three long DNA fragments (ODNs 6, 7 and 8), the central one (ODN 7) being functionalised with both 5'-amino and 3'-phosphate. Complementary splints ODN 9 and 10 were used to template the reaction. Variations on the composition of the ligation buffer used in synthesis of ODN 1 were evaluated and 0.4 M sodium chloride and 0.2 M HEPES at pH 7.2 were found to be most suitable (SI Figure S4). Lower concentrations of sodium chloride are not as effective. The reaction is not strongly sensitive to pH in the range 7.0 to 8.5.

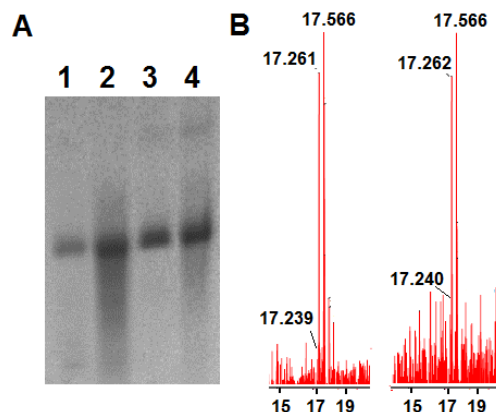
The 303-mer DNA ligation product was amplified by PCR, cloned into a sequencing vector, and 20 individual clones were sequenced, all of which displayed the correct nucleobase sequence at the ligation sites (SI Figure S6). The majority of the sequences were free of mutations (13 out of 20 clones), but in some strands a small number of mutations were observed, located some distance from the phosphoramidate linkages. We identified 12 mutations in total, of which 10 were deletions, one was an insertion and only one was a point mutation (G for A) with zero mutations around the ligation sites. This mutation rate is in line with that expected from standard oligonucleotide synthesis. Importantly it is unlikely

that any of the mutations could have originated from side reactions due to the coupling agent. Finally, the full 762 base pair gene for green fluorescent protein (EGFP) was synthesised in a single tube from 10 overlapping amine/phosphate functionalised ODNs ranging from 120 to 177 bases in length (ODN 15 to ODN 24). PCR amplification followed by cloning and Sanger sequencing indicated that the full length product was obtained, with only one deletion mutation which was far



from the ligation point, (Figure 2 and SI Figure S8).

Figure 2. A) PCR amplification of the double stranded phosphoramidate EGFP gene (762-mer). Lane 1; 100 bp DNA ladder, lane 2; PCR using the double strand phosphoramidate EGFP gene (762-mer), lane 3; control PCR for individual oligos without ligation. B) Representative sequencing data from cloning of the PCR product of the phosphoramidate EGFP gene. The data show that the polymerase faithfully copied the bases around the phosphoramidate ligation sites (shown in



red in the sequence text).

Figure 3. A) Transcription of 79-mer phosphoramidate and DNA unmodified templates. Lanes 1 and 2; reaction using Phosphoramidate template (ODN 27) and long coding strand (ODN 32) for 2h and 4 h respectively, lanes 3 and 4; reaction using control template (ODN 31) and long coding strand (ODN 32) for 2h and 4 h respectively, 15% Polyacrylamide gel. B) ES- mass spectra of the RNA transcripts from the phosphoramidate template (left) and the normal unmodified template (ODN 31) (right). The transcripts both have the expected 5'-triphosphate. In addition to the full-length transcript peak, another peak was observed in both cases corresponding to the full length construct + 3'-cytidine.

Having established that gene synthesis by chemical ligation is possible we turned our attention to transcription. Using the phosphoramidate ligation methodology we constructed a 79-mer DNA template containing a single phosphoramidate linkage which was used as a template to synthesise a 53-mer RNA strand. Transcription was carried out from a fully complementary duplex or the single stranded template with a double stranded T7 promotor region. Both transcription reactions were successful and, as expected, the former was more efficient. The transcription product was characterised by

gel electrophoresis and mass spectrometry (Figure 3 and SI Figure S11 and S12).

The orthogonality of phosphoramidate and CuAAC copper click ligation could be important in many applications, for example DNA encoded library assembly, where switching between ligation chemistries could allow a greater variety of chemical reactions to be used in assembling small molecule libraries. We demonstrated this orthogonality by simultaneously ligating three oligonucleotides ODN 6, ODN 37 and ODN 38 in a single tube, the central one (ODN 38) having a 5'-amine and 3'-alkyne, to create a 303-mer DNA strand containing one biocompatible phosphoramidate linkage and one biocompatible triazole linkage<sup>5</sup> (Figure 4, Figure S9). An additional experiment was carried out to make a fluorescent 331-mer DNA construct and thereby demonstrate the generality of the method (Figure S10).

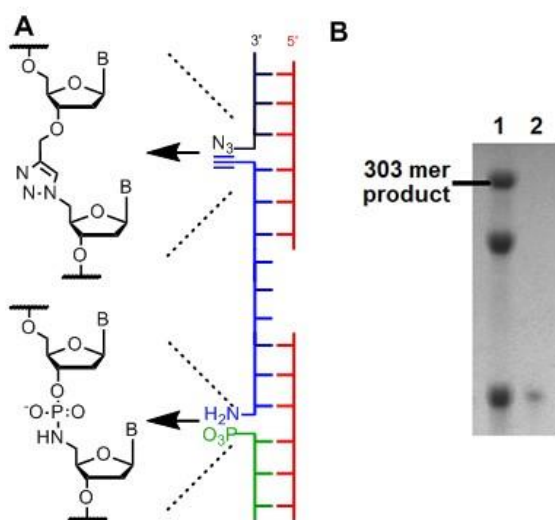


Figure 4. Orthogonal phosphoramidate and CuAAC reactions for ligation of three oligonucleotides to make a 303-mer product. Lane 1; orthogonal phosphoramidate and CuAAC reactions using ODN 6 (3'-phosphate), ODN 38 (3'-propargyl and 5'-amine) and ODN 37 (5'-azide), lane 2; starting material ODN 6. Denaturing 8% polyacrylamide gel-electrophoresis.

In conclusion we have demonstrated that 5'-amino and 3'-phosphate oligonucleotides can be used in multiple sequential templated DNA ligation reactions. The resultant 5'-phosphoramidate linkage can be read through efficiently by a thermostable DNA polymerase and by T7 RNA polymerase. The EDC coupling agent does not react in an undesirable manner with DNA and therefore is not mutagenic under chemical ligation conditions. The scope of this gene synthesis methodology has been demonstrated by construction of a 762 base pair gene that contains 8 phosphoramidate linkages. The amine and phosphate-modified oligonucleotide building blocks are readily synthesised using established methods and the approach should be suitable for high-throughput automation and the production of genes on a large scale. This straightforward methodology could be readily adopted by commercial oligonucleotide production and gene synthesis facilities. It greatly increases the scope of chemistry-based DNA synthesis, being suitable for the assembly of very long DNA strands that contain sugar, nucleobase, and various backbone modifications. Furthermore, it can be used

simultaneously with CuAAC ligation to produce nucleic acids containing mixed backbone analogues.

## Acknowledgment

This work was supported by UK BBSRC sLoLa grant BB/J001694/2: Extending the boundaries of nucleic acid chemistry and UK BBSRC grants BB/M025624/1 (Next-generation DNA synthesis).

## Notes and references

1. S. Kosuri and G. M. Church, *Nature Meth.*, 2014, 11, 499-507.
2. M. M. Suzuki and A. Bird, *Nature Rev. Genet.*, 2008, 9, 465-476.
3. S. de Kok, L. H. Stanton, T. Slaby, M. Durot, V. F. Holmes, K. G. Patel, D. Platt, E. B. Shapland, Z. Serber, J. Dean, J. D. Newman and S. S. Chandran, *ACS Syn. Biol.*, 2014, 3, 97-106.
4. A. S. Xiong, R. H. Peng, J. Zhuang, J. G. Liu, F. Gao, J. M. Chen, Z. M. Cheng and Q. H. Yao, *Biotechnol. Adv.*, 2008, 26, 121-134.
5. A. H. El-Sagheer, A. P. Sanzone, R. Gao, A. Tavassoli and T. Brown, *Proc. Natl. Acad. Sci. USA.*, 2011, 108, 11338-11343.
6. A. Shivalingam, A. E. S. Tyburn, A. H. El-Sagheer and T. Brown, *J. Amer. Chem. Soc.*, 2017, in press, 10.1021/jacs.1026b11530.
7. A. H. El-Sagheer and T. Brown, *Proc. Natl. Acad. Sci. USA.*, 2010, 107, 15329-15334.
8. J. A. Doudna and E. Charpentier, *Science*, 2014, 346, 12580961 - 12580969.
9. C. N. Birts, A. P. Sanzone, A. H. El-Sagheer, J. P. Blaydes, T. Brown and A. Tavassoli, *Angew. Chem. Int. Ed.*, 2014, 53, 2362-2365.
10. X. Chen, A. H. El-Sagheer and T. Brown, *ChemComm*, 2014, 50, 7597-7600.
11. A. H. El-Sagheer and T. Brown, *ChemComm*, 2011, 47, 12057-12058.
12. S. Peyrottes, J.-J. Vasseur, J.-L. Imbach and B. Rayner, *Nucleic Acids Research*, 1996, 24, 1841-1848.
13. S. M. Gryaznov and R. L. Letsinger, *Nucleic Acids Research*, 1992, 20, 3403-3409.
14. M. Kalinowski, R. Haug, H. Said, S. Piasecka, M. Kramer and C. Richert, *Chembiochem*, 2016, 17, 1150-1155.