Synthesis of chemically modified DNA

Arun Shivalingam* and Tom Brown*1

*Department of Chemistry, University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford, OX1 3TA, U.K.

Abstract

Naturally occurring DNA is encoded by the four nucleobases adenine, cytosine, guanine and thymine. Yet minor chemical modifications to these bases, such as methylation, can significantly alter DNA function, and more drastic changes, such as replacement with unnatural base pairs, could expand its function. In order to realize the full potential of DNA in therapeutic and synthetic biology applications, our ability to 'write' long modified DNA in a controlled manner must be improved. This review highlights methods currently used for the synthesis of moderately long chemically modified nucleic acids (up to 1000 bp), their limitations and areas for future expansion.

Introduction

The ordered and addressable arrangement of the four nucleobases adenine (A), cytosine (C), guanine (G) and thymine (T) in DNA allows our cells to store approximately 1–2 GB of data in 7 pg of material. The accessibility of these data is carefully controlled via chemical modification of the nucleobases, double helical DNA formation and compaction, as well as numerous DNA–protein interactions. Of these mechanisms, chemical modifications are of particular interest in that they increase the information storage capacity of DNA and provide functional handles for external manipulation. For example, controlled DNA methyltransferase-mediated methylation of cytosine modulates gene expression [1], whereas uncontrolled reactive oxygen species can oxidize nucleobases inducing mutations upon polymerase replication. Top-down studies have developed our understanding of these processes, however our ability to exploit this knowledge in synthetic biology applications and expand upon it using bottom-up approaches for therapeutic applications is constrained by our inability to 'read and write' modified DNA efficiently.

In the past decade, remarkable progress has been made in both of the above aspects for unmodified nucleic acids. Millions of DNA strands can now be sequenced in parallel using Illumina and Ion Torrent next generation sequencing platforms [2], whereas whole genomes such as that of *Mycoplasma mycoides* can be reconstructed from oligonucleotides [3]. Yet of the two processes, our ability to write still lags behind our ability to read nucleic acid sequences. Single base resolution of modified nucleobases, such as 5-methyl-, hydroxy-, formyl- and carboxy-C, is now possible [4], and with the rapid development of single molecule polymerase-independent nanopore sequencing this number may expand considerably still [5]. On the other hand, the *in vitro* synthesis of highly ordered modified

DNA, whether it contains (for example) amino acid side chains, epigenetic modifications, electrochemical markers or fluorophore tags, is still in its infancy partly due to the reliance of gene synthesis clean-up and amplification by PCR. Consequently, the focus of this review is to evaluate the current state of modified DNA synthesis (up to 1000 bp) and highlight the challenges that have yet to be overcome.

Chemical synthesis

Automated solid-phase phosphoramidite-based synthesis allows the routine and cheap production of oligonucleotides of approximately 100 bases. The principal chemical steps involved in each cycle were optimized by the early 1990s and have remained largely unchanged since (Figure 1) [6,7]. Stepwise coupling of phosphoramidites allows unparalleled control of DNA modification, however this comes at the cost of imperfect coupling (approximately 98.5–99.5 %) and uncontrolled mutagenesis due to chemical exposure [8]. Furthermore, the modifications to be introduced must be compatible with mildly acidic, strongly basic and oxidizing conditions used in the various synthetic steps. This can be particularly challenging when multiple modifications, each with their own unique deprotection requirements, are used. When combined, this limits the length and yield of the oligonucleotide (if coupling efficiency=99 %, 100-mer maximum yield=36.6 %) as well as the number and type of modifications that can incorporated (Figure 2).

Efforts to optimize each of the four steps in the phosphoramidite oligonucleotide synthesis cycle have resulted in a diverse range of reagents and conditions. For example, tetrazole, benzylthiotetrazole (BTT) or ethylthiotetrazole (ETT) coupling reagents could be used, each with their own optimal concentrations, incubation times and cycle repetitions; protecting groups could be removed in concentrated ammonium hydroxide, 0.1 M NaOH(aq) or methanolic potassium carbonate solution at temperatures ranging from 25 to 55 °C for 1–18 h [9]. Yet the gains in purity and yield due to changes in these specific parameters are challenging to

Figure 1 | Standard automated solid-phase phosphoramidite-based oligonucleotide synthesis

The process consists of four steps (standard conditions in brackets) – (1) 5'-hydroxy detritylation (TCA/CH$_2$Cl$_2$), (2) phosphoramidite coupling (monomer, tetrazole), (3) uncoupled 5'-hydroxy capping (Ac$_2$O/pyridine/THF, N-methylimidazole in MeCN) and (4) phosphorus oxidation (I$_2$, H$_2$O/Pyridine/THF).
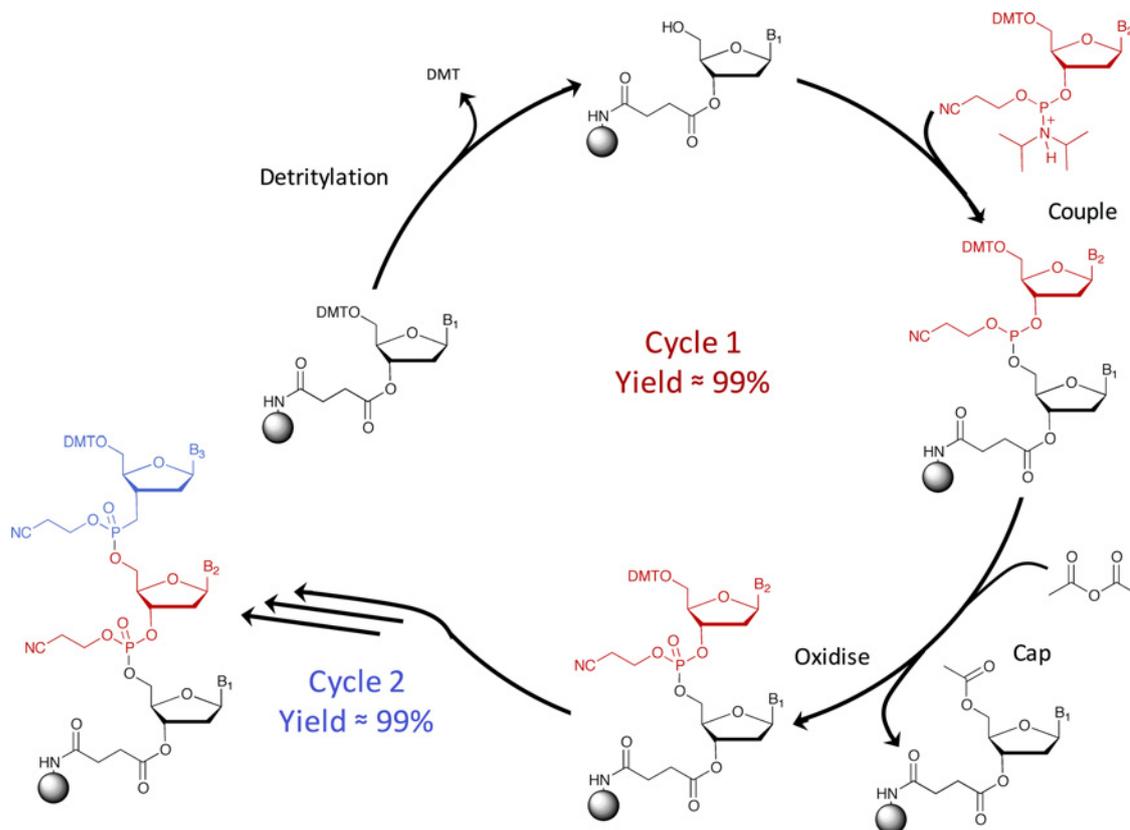
quantify since they are likely to be interdependent and result in deletions, insertions or point mutations. Until recently [10], this have been difficult to characterize in an accurate and cost-effective manner.

On the other hand, evaluation of depurination or incomplete detritylation, both of which impede the synthesis of full-length oligonucleotides and are linked to acidity, is more straightforward due to techniques for size separation of oligonucleotides and spectroscopic analysis. Alternative detritylation approaches include replacement of the Brønsted acid trichloroacetic acid (TCA) with chelating Lewis acids such as zinc bromide or mildly acidic sodium acetate buffer (10 mM, pH 3.5) [11,12]. Unfortunately, the former requires use of the explosive nitromethane for rapid detritylation and the latter takes approximately 30 min for near quantitative conversion, reducing their viability in commercial applications. Depurination-inhibiting protecting groups such as amidines that decrease the pK_a of the N7 (or N1) position of purines have also been explored but have not been widely adopted possibly due to slower deprotection rates [13]. Finally, with the advent of high-throughput microarray oligonucleotide synthesizers with capacity for parallel synthesis of 12000–1000000

oligonucleotides, variations in platform design have given rise to the largest innovations in detritylation chemistry. Electrochemically generated acids allow deprotection in approximately 5 s [14], whereas photolabile 5'- protecting groups avoid the use of acid altogether [15]. For inkjet printed microarrays, it was found that active quenching of the acid by the oxidizing solution significantly reduced depurination [16]. However, these gains are tempered by losses in spatial resolution of reagents due to high-throughput synthesis compared with conventional column synthesis, which ultimately constrains the maximum length of modified oligonucleotide synthesis to approximately 150–200 bases.

Polymerase-based synthesis

Therefore, the production of larger modified DNA must rely upon the assembly of smaller components. In this regard, the simplest method is to create the desired sequence by PCR amplification methods using a modified dNTP in place of the naturally occurring one. A key requirement for such dNTPs is that they must be incorporated and also read by polymerases. As a result, modifications that do not disturb the Watson–Crick base pairing face of the nucleobases are

Figure 2 | Examples of base modified phosphoramidities, triphosphates and unnatural base pairs

The position of the modification is expressed relative to the numbering system for used for nucleic acids. For phosphoramidites, protecting groups are omitted for structural clarity. More comprehensive lists of phosphoramidites are available from commercial suppliers such as Link Technologies, Glen Research, Berry Associates, Chemgenes and LGC Biosearch Technologies. For triphosphates, a set of simpler C5 pyrimidine modifications are shown that highlight the various functional handles that can be introduced and used for subsequent manipulation. For unnatural bp, the three most successful examples reported to date are shown. For dPx, R=H or CH(OH)CH$_2$(OH).
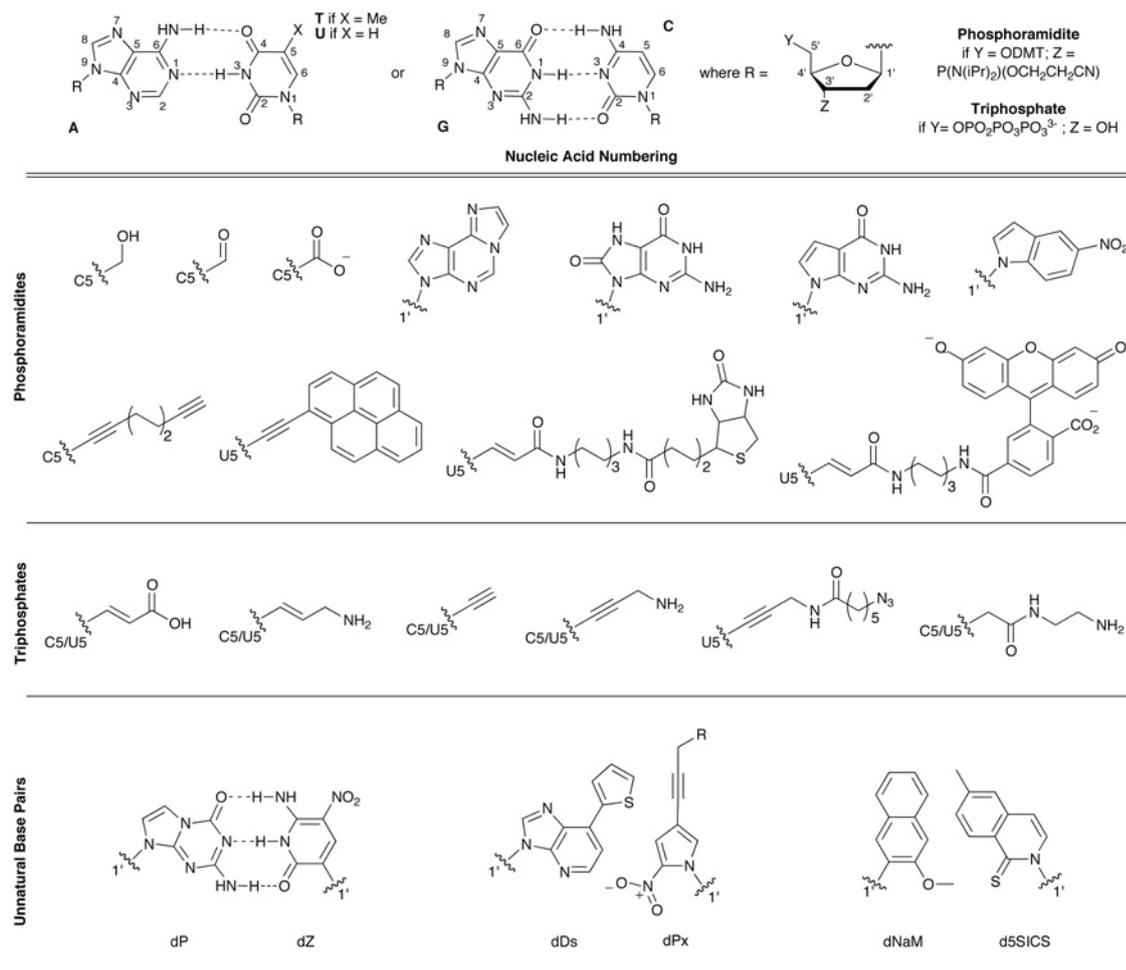
necessary; the C5 position of pyrimidines and the C7 position of 7-deazapurines are usually chosen for functionalization due to protrusion of the modification into the major groove of DNA, a structural feature that is more readily tolerated by polymerases (based on biochemical experiments and X-ray crystallographic data) [17,18]. Of these two positions, the C5 pyrimidine modifications are more synthetically accessible and consequently more widely reported (Figure 2) [19,20]. C8 purine modifications have been reported but their incorporation efficiency is noticeably worse [21,22]. It should also be noted that the (de)stabilizing properties of the modification must be carefully modulated for very large DNA duplex synthesis since most modifications, with the exception of C5-alkynyl pyrimidines and C7-alkynl 7-deazapurines [23,24], disrupt DNA duplex thermal stability by up to 1–2 °C per incorporation.

Correct optimization of PCR extension time (typically lengthened) and temperature (typically lowered) is important for modified DNA synthesis but the choice of polymerase is most critical. Family B polymerases, such as *KOD XL*, *Vent (exo-)* and *Pwo*, are generally superior to family A polymerases [19,25]. Indeed, for Cy3 and Cy5 fluorophore labelled dNTPs, genetically evolved family B *Pfu* variants are required for full incorporation of the modification by PCR [26]. The properties of the resultant 1000 bp product neatly illustrates the fascinating behaviour of heavily modified DNA, in that conventional small molecule DNA intercalation was no longer possible and solubility was higher in organic rather than aqueous solvents. The latter point does, however, highlight that it would be favourable to control the density of modifications, which is sometimes achieved through the use of a mixture of modified and unmodified

dNTPs [27]. This is most evident when bulky modifications must be incorporated sequentially along a template; this results in some templates being more 'modifiable' than other.

Perhaps the main limitation of this method is that it is an 'all or nothing' approach. If all A, C, G or T sites of the product are replaced by a modified nucleotide (X), then the product is a defined single species; if there is a mixture of modifications introduced via a mixture of a natural dNTP and the unnatural dXTP for either of the A, C, G or T sites, then a range of products is obtained with varying modification content. For some applications, such as synthesis of fluorescent *in situ* hybridization probes, the mixed products could be of benefit as this enables custom mixtures of multiple fluorophores that give rise to unique optical output [27,28]. Indeed, for aptamer (nucleic acid antibody equivalents) selection, this would be ideal as fully modified sequences may not have sufficient structural diversity for target binding [24]. Yet the problem lies in identifying the position of the modified base and, if more than one round of aptamer selection is required, its site-specific re-introduction into an enriched subpopulation of sequences.

Finally, if single PCR products are required, given that polymerases have evolved to recognize only the naturally occurring bases/bp (AT, TA, CG, GC), only four modifications can be made per PCR generated product [25]. This premise has been challenged in recent years through the use of unnatural bp. To date three unnatural bp have been reported with very high stability and replication fidelity (>99.8%) under defined conditions (Figure 2) [29–31]. Selectivity is maintained either by alternative hydrogen bonding patterns or predominantly hydrophobic interactions. The 'expanded genetic alphabet' is particularly intriguing due its promise of enhanced data storage but has so far been used mainly for aptamer selection where its presence was vital for function of the identified aptamer; substitution of the modified base for its natural base analogue resulted in a >100-fold decrease in affinity of the aptamer for the target [32,33]. This effect is most likely due to the unnatural hydrophobic interactions these bases generate as opposed to the expanded sequence space. Although these studies did not demonstrate that control unmodified sequences fail to generate aptamers of similar affinity, the importance of hydrophobic 'protein-like' interactions has been established for C5-modified pyrimidine sequences against a range of protein targets [34]. Although the dNaM–d5SICS bp has been shown to be biocompatible in *Escherichia coli*, caution must be taken as the dNTPs required for its replication cannot be synthesized natively in cells [35].

Ligation-based synthesis

The above PCR-based methods cannot be used to produce chemically modified DNA containing specific modifications at predefined loci. For such site-specific incorporation of modifications in long DNA, the ideal reaction would involve direct stepwise ligation of shorter chemically synthesized modified oligonucleotides. This would bypass the inability

of polymerases to detect modifications that are not on present the Watson–Crick hydrogen bonding face, and would allow for a much wider range of modifications, including those to sugar and phosphate. To this end, DNA ligases can be used, however the templated assembly of >1000 bp genes using this method has been restricted to unmodified oligonucleotides [36–38]. This is surprising given the growing interest in the epigenome and in particular the role that oxidized forms of 5-methyl-C play in it [39]. A reason for this could be ligation inefficiency; post-ligation enrichment of the desired gene by PCR or cloning is often used [38], which would erase all modifications introduced. Yet, with optimization, ligation methods could feasibly be used to synthesize large DNA constructs with highly ordered and predetermined arrangements of modifications. Precedent for this comes from templated assembly of 5' phosphorylated trimers and pentamers bearing peptide moieties using T4 DNA ligase [40,41]. Products of 50–150 bases were synthesized (10–50 independent ligations) with efficiencies reaching approximately 98%. Promisingly, these products could be read by *Vent (exo-)* polymerase and re-synthesized by T4 DNA ligase, enabling their use in inhibitor identification by *in vitro* selection.

In this context, chemical ligation is appealing in that small molecule chemical reactions can be highly efficient, and should be more agnostic than ligase enzymes in terms of the substrates to be ligated. The first example of this was a condensation reaction using cyanogen bromide [42], whereas later methods involved the spontaneous displacement of a tosylate/iodo group from the 5'- position of oligonucleotides using 3' thiophosphorylated oligonucleotides [43,44]. Although promising, these methods failed to gain wider adoption due to the lack of controlled ligation, stability of components in water and toxicity. To this end, development of bifunctional 5'-azide 3'-alkyne modified oligonucleotides for use in copper-catalysed cycloaddition 'click' chemistry was developed in our group (Figure 3) [45,46]. The resultant triazole linkage, although not a natural phosphodiester, is remarkably biocompatible and non-toxic. Successive refinement of the structure based on the hypothesis that the triazole acts as a hydrogen bond acceptor, which was later supported by NMR studies [47], in combination with efforts to reduce the synthetic burden of monomer synthesis resulted in a second generation version that is faithfully replicated [48]. This is the first fully modified nucleic acid backbone that is compatible with DNA and RNA polymerases *in vitro* and in mammalian cells [49]. As proof of principle, a 300-base oligonucleotide containing two triazole linkages was prepared. Further work is ongoing to adapt this methodology for larger modified DNA synthesis.

Post-synthetic modification

In principle, all of the above methodologies can be used in conjunction with post-synthetic modification of DNA. In effect, this reduces the chemical or enzymatic

Figure 3 | Unnatural biocompatible 'click' ligation of oligonucleotides

Chemical ligation using copper-catalysed azide-alkyne cycloaddition (CuAAC) 'click' chemistry to generate first and second generation biocompatible triazole internucleoside linkages.
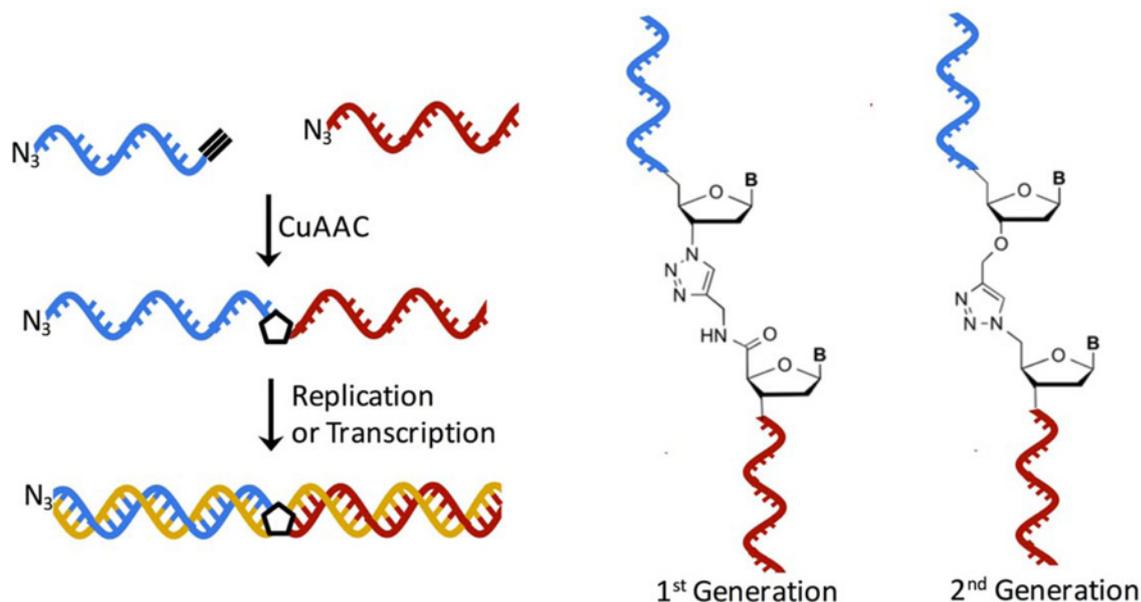
synthetic burden by reducing monomer complexity. Simple small functional handles, such as aldehydes, linear alkynes, cycloalkynes, azides, carboxylic acids or primary amines, can be introduced as nucleobase modifications during chemical or enzymatic synthesis as phosphoramidites and triphosphates, before being subsequently reacted to generate the final desired modification. These handles could also be generated *de novo* using methyltransferases; requiring a recognition motif typically up to 6 bp in length, these enzymes have been used to transfer moieties larger than methyl groups through the use of modified *S*-adenosyl methionine substrates [50,51]. A second pre-labelling step is sometimes required to access handles such as azides due to the chemical instability during solid-phase oligonucleotide synthesis. Of the numerous coupling methods that have been established, there are to date just a few orthogonal reactions [52,53] (e.g. amide formation, inverse Diels–Alder reaction [54], copper-catalysed [55] /strain-promoted [56] Huisgen [2 + 3] cycloaddition and oxime formation [57]). As a consequence, selective protection of different moieties is required for the introduction of multiple modifications, with one of the best examples being the two different levels of terminal alkyne protection used for three independent copper 'click' reactions by solid-phase synthesis [58]. Despite its synthetic ease, the main drawbacks to post-synthetic functionalization are incomplete or slow reactions depending upon reactant concentrations, partly limited by solubility issues. Hence, the methodology is most efficiently used with solid-phase immobilized oligonucleotides, where the reactants can be used in excess in a suitable solvent and easily removed.

Conclusions

The synthesis of highly modified DNA is constrained by the limited ability of polymerases to recognize modifications to the nucleobase, phosphate or sugar. Consequently, the chemical complexity of polymerase generated DNA will always be limited by the number of different dNTPs that are available and the number of different bases or bp that can be read and written by the polymerase enzyme. If locus/sequence specific modifications to DNA are required, the problem becomes intractable. On the other hand, for the generation of random libraries, the use of modified dNTPs has clear advantages but is currently handicapped by our limited ability to decode the modifications and the low/inconsistent yields of triphosphate (dNTP) synthesis [59]; novel high yielding methodologies are required. For ordered and patterned synthesis of modified DNA, it is evident that a combination of chemical and enzymatic synthetic methods should be used. The high number of coupling steps achievable by solid-phase synthesis (up to 150 with appreciable final yields) highlights that chemical coupling can be exceptionally efficient, but that it does have a finite limit. This is partly due to the architecture of the solid supports used conventionally. Therefore, to achieve long modified DNA synthesis, a compromise must be made between oligonucleotide length, purity and the number of ligation reactions. The issue of oligonucleotide purity is particularly important given that post-assembly error correction methods currently used in PCR-based gene synthesis [60] will have detrimental effects when synthesizing modified DNA; they can remove modifications and replace

them with unmodified sections of DNA. Therefore, it is essential that solid-phase automated oligonucleotide synthesis, in addition to ligation methods, be optimized if the synthesis of very long chemically modified DNA constructs is to become a routine exercise in the future.

Funding

References

1 Song, J., Rechkoblit, O., Bestor, T.H. and Patel, D.J. (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science **331**, 1036–1040 CrossRef PubMed

2 Metzker, M.L. (2010) Sequencing technologies – the next generation. Nat. Rev. Genet. **11**, 31–46 CrossRef PubMed

3 Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.,Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M. et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science **329**, 52–56 CrossRef PubMed

4 Booth, M.J., Raiber, E.,A. and Balasubramanian, S. (2015) Chemical methods for decoding cytosine modifications in DNA. Chem. Rev. **115**, 2240–2254 CrossRef PubMed

5 Schreiber, J., Wescoe, Z.L., Abu-Shumays, R., Vivian, J.T., Baatar, B., Karplus, K. and Akeson, M. (2013) Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. Proc. Natl. Acad. Sci. U.S.A. **110**, 18910–18915 CrossRef PubMed

6 Caruthers, M.H., Barone, A.D., Beaucage, S.L., Dodds, D.R., Fisher, E.F., McBride, L.J., Matteucci, M., Stabinsky, Z. and Tang, J.-Y. (1987) Chemical synthesis of deoxyoligonucleotides by the phosphoramidite method. Methods Enzymol. **154**, 287–313 CrossRef

7 Caruthers, MH. (1991) Chemical synthesis of DNA and DNA analogs. Acc. Chem. Res. **24**, 278–284 CrossRef

8 Rayner, S., Brignac, S., Bumeister, R., Belosludtsev, Y., Ward, T., Grant, O., O'Brien, K., Evans, G.A. and Garner, H.R. (1998) MerMade: an oligodeoxyribonucleotide synthesizer for high throughput oligonucleotide production in dual 96-well plates. Genome. Res. **8**, 741–747 PubMed

9 Ellington, A. and Pollard, J.D. (2001) Introduction to the synthesis and purification of oligonucleotides. Current Protocols in Nucleic Acid Chemistry. John Wiley & Sons, Oxford

10 Schwartz, J.J., Lee, C. and Shendure, J. (2012) Accurate gene synthesis with tag-directed retrieval of sequence-verified DNA molecules. Nat. Meth. **9**, 913–915 CrossRef

11 Krotz, A.H., Cole, D.L. and Ravikumar, V.T. (2003) Synthesis of antisense oligonucleotides with minimum depurination. Nucleosides, Nucleotides Nucleic Acids **22**, 129–134 CrossRef PubMed

12 Matteucci, M.D. and Caruthers, M.H. (1980) The use of zinc bromide for removal of dimethoxytrityl ethers from deoxynucleosides. Tetrahedron Lett. **21**, 3243–3246 CrossRef

13 Froehler, B.C. and Matteucci, M.D. (1983) Dialkylformamidines: depurination resistant N6-protecting group for deoxyadenosine. Nucleic Acids Res. **11**, 8031–8036 CrossRef PubMed

14 Egeland, R.D. and Southern, E.M. (2005) Electrochemically directed synthesis of oligonucleotides for DNA microarray fabrication. Nucleic Acids Res. **33**, e125–e125 CrossRef PubMed

15 Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P. and Fodor, S.P. (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc. Natl. Acad. Sci. U.S.A. **91**, 5022–5026 CrossRef PubMed

16 LeProust, E.M., Peck, B.J., Spirin, K., McCuen, H.B., Moore, B., Namsaraev, E. and Caruthers, M.H. (2010) Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. Nucleic Acids Res. **38**, 2522–2540 CrossRef PubMed

17 Obeid, S., Baccaro, A., Welte, W., Diederichs, K. and Marx, A. (2010) Structural basis for the synthesis of nucleobase modified DNA by Thermus aquaticus DNA polymerase. Proc. Natl. Acad. Sci. U.S.A. **107**, 21327–21331 CrossRef PubMed

18 Bergen, K., Steck, A.-L., Strütt, S., Baccaro, A., Welte, W., Diederichs, K. and Marx, A. (2012) Structures of KlenTaq DNA polymerase caught while incorporating C5-modified pyrimidine and C7-modified 7-deazapurine nucleoside triphosphates. J. Am. Chem. Soc. **134**, 11840–11843 CrossRef PubMed

19 Kuwahara, M., Nagashima, J., Hasegawa, M., Tamura, T., Kitagata, R., Hanawa, K., Hososhima, S., Kasamatsu, T., Ozaki, H. and Sawai, H. (2006) Systematic characterization of 2′-deoxynucleoside- 5′-triphosphate analogs as substrates for DNA polymerases by polymerase chain reaction and kinetic studies on enzymatic production of modified DNA. Nucleic Acids Res. **34**, 5383–5394 CrossRef PubMed

20 Ren, X., El-Sagheer, A.H. and Brown, T. (2015) Azide and trans-cyclooctene dUTPs: incorporation into DNA probes and fluorescent click-labelling. Analyst **140**, 2671–2678 CrossRef PubMed

21 Thomas, J.M., Yoon, J.-K. and Perrin, D.M. (2009) Investigation of the catalytic mechanism of a synthetic DNAzyme with protein-like functionality: An RNaseA mimic? J. Am. Chem. Soc. **131**, 5648–5658 CrossRef

22 Čapek, P., Cahová, H., Pohl, R., Hocek, M., Gloeckner, C. and Marx, A. (2007) An efficient method for the construction of functionalized DNA bearing amino acid groups through cross-coupling reactions of nucleoside triphosphates followed by primer extension or PCR. Chem. Eur. J. **13**, 6196–6203 CrossRef

23 Seela, F., Budow, S., Shaikh, K.I. and Jawalekar, A.M. (2005) Stabilization of tandem dG-dA base pairs in DNA-hairpins: replacement of the canonical bases by 7-deaza-7-propynylpurines. Org. Biomol. Chem. **3**, 4221–4226 CrossRef PubMed

24 Kraemer, S., Vaught, J.D., Bock, C., Gold, L., Katilius, E., Keeney, T.R., Kim, N., Saccomano, N.A., Wilcox, S.K., Zichi, D. et al. (2011) From SOMAmer-based biomarker discovery to diagnostic and clinical applications: a SOMAmer-based, streamlined multiplex proteomic assay. PLoS One **6**, e26332 CrossRef PubMed

25 Jäger, S., Rasched, G., Kornreich-Leshem, H., Engeser, M., Thum, O. and Famulok, M. (2005) A versatile toolbox for variable DNA functionalization at high density. J. Am. Chem. Soc. **127**, 15071–15082 CrossRef PubMed

26 Ramsay, N., Jemth, A.S, Brown, A., Crampton, N., Dear, P. and Holliger, P. (2010) CyDNA: synthesis and replication of highly Cy-dye substituted DNA by an evolved polymerase. J. Am. Chem. Soc. **132**, 5096–5104 CrossRef PubMed

27 Tasara, T., Angerer, B., Damond, M., Winter, H., Dörhöfer, S., Hübscher, U. and Amacker, M. (2003) Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. II. High-density labeling of natural DNA. Nucleic Acids Res. **31**, 2636–2646 CrossRef PubMed

28 Ren, X., El-Sagheer, A.H. and Brown, T. (2016) Efficient enzymatic synthesis and dual-colour fluorescent labelling of DNA probes using long chain azido-dUTP and BCN dyes. Nucleic Acids Res., http://nar.oxfordjournals.org/content/early/2016/01/26/nar.gkw028.abstract

29 Kimoto, M., Kawai, R., Mitsui, T., Yokoyama, S. and Hirao, I. (2009) An unnatural base pair system for efficient PCR amplification and functionalization of DNA molecules. Nucleic Acids Res. **37**, e14–e14 CrossRef PubMed

30 Malyshev, D.A., Dhami, K., Quach, H.T., Lavergne, T., Ordoukhanian, P., Torkamani, A. and Romesberg, F.E. (2012) Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. Proc. Natl. Acad. Sci. U.S.A. **109**, 12005–12010 CrossRef PubMed

31 Yang, Z., Chen, F., Alvarado, J.B. and Benner, S.A. (2011) Amplification, mutation, and sequencing of a six-letter synthetic genetic system. J. Am. Chem. Soc. **133**, 15105–15112 CrossRef PubMed

32 Sefah, K., Yang, Z., Bradley, K.M., Hoshika, S., Jiménez, E., Zhang, L., Zhu, G., Shanker, S., Yu, F., Turek, D. et al. (2014) *In vitro* selection with artificial expanded genetic information systems. Proc. Natl. Acad. Sci. U.S.A. **111**, 1449–1454 CrossRef PubMed

33 Kimoto, M., Yamashige, R., Matsunaga, K., Yokoyama, S. and Hirao, I. (2013) Generation of high-affinity DNA aptamers using an expanded genetic alphabet. Nat. Biotech. **31**, 453–457 CrossRef

34 Gold, L., Ayers, D., Bertino, J., Bock, C., Bock, A., Brody, E.N., Carter, J., Dalby, A.B., Eaton, B.E., Fitzwater, T. et al. (2010) Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One **5**, e15004 CrossRef PubMed

35 Malyshev, D.A., Dhami, K., Lavergne, T., Chen, T., Dai, N., Foster, J.M., Correa, I.R. and Romesberg, F.E. (2014) A semi-synthetic organism with an expanded genetic alphabet. Nature **509**, 385–388 CrossRef PubMed

36 Parker, H.Y. and Mulligan, J. (2003) Solid phase methods for polynucleotide production. US20030228602.

37 Schatz, O., O'connell, T, Schwer, H. and Waldmann, T. (2008), Method for the manufacture of nucleic acid molecules. EP1411122

38 Au, L.C., Yang, F.Y., Yang, W.J., Lo, S.H. and Kao, C.F. (1998) Gene synthesis by a LCR-based approach: High-level production of Leptin-L54 using synthetic gene in *Escherichia coli*. Biochem. Biophys. Res. Commun. **248**, 200–203 CrossRef PubMed

39 Wu, H. and Zhang, Y. (2015) Charting oxidized methylcytosines at base resolution. Nat. Struct. Mol. Biol. **22**, 656–661 CrossRef PubMed

40 Guo, C., Watkins, C.P. and Hili, R. (2015) Sequence-defined scaffolding of peptides on nucleic acid polymers. J. Am. Chem. Soc. **137**, 11191–11196 CrossRef PubMed

41 Hili, R., Niu, J. and Liu, D.R. (2013) DNA ligase-mediated translation of DNA into densely functionalized nucleic acid polymers. J. Am. Chem. Soc. **135**, 98–101 CrossRef PubMed

42 Shabarova, Z.A., Merenkova, I.N., Oretskaya, T.S., Sokolova, N.I., Skripkin, E.A., Alexeyeva, E.V., Balakin, A.G. and Bogdanov, A.A. (1991) Chemical ligation of DNA: the first non-enzymatic assembly of a biologically active gene. Nucleic Acids Res. **19**, 4247–4251 CrossRef PubMed

43 Xu, Y. and Kool, E.T. (1998) Chemical and enzymatic properties of bridging 5'-S-phosphorothioester linkages in DNA. Nucleic Acids Res. **26**, 3159–3164 CrossRef PubMed

44 Herrlein, M.K., Nelson, J.S. and Letsinger, R.L. (1995) A covalent lock for self-assembled oligonucleotide conjugates. J. Am. Chem. Soc. **117**, 10151–10152 CrossRef

45 El-Sagheer, A.H. and Brown, T. (2009) Synthesis and polymerase chain reaction amplification of DNA strands containing an unnatural triazole linkage. J. Am. Chem. Soc. **131**, 3958–3964 CrossRef PubMed

46 Qiu, J., El-Sagheer, A.H. and Brown, T. (2013) Solid phase click ligation for the synthesis of very long oligonucleotides. Chem. Commun. **49**, 6959–6961 CrossRef

47 Dallmann, A., El-Sagheer, A.H., Dehmel, L., Mügge, C., Griesinger, C., Ernsting, N.P. and Brown, T. (2011) Structure and dynamics of triazole-linked DNA: Biocompatibility explained. Chem. Eur. J. **17**, 14714–14717 CrossRef

48 El-Sagheer, A.H., Sanzone, A.P., Gao, R., Tavassoli, A. and Brown, T. (2011) Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **108**, 11338–11343 CrossRef PubMed

49 Birts, C.N., Sanzone, A.P., El-Sagheer, A.H., Blaydes, J.P., Brown, T. and Tavassoli, A. (2014) Transcription of click-linked DNA in human cells. Angew. Chem. Int. Ed. **53**, 2362–2365 CrossRef

50 Zhang, J. and Zheng, Y.G. (2016) SAM/SAH analogs as versatile tools for SAM-dependent methyltransferases. ACS Chem. Biol. **11**, 583–597 CrossRef PubMed

51 Dalhoff, C., Lukinavicius, G., Klimasauskas, S. and Weinhold, E. (2006) Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. Nat. Chem. Biol. **2**, 31–32 CrossRef PubMed

52 Gutsmiedl, K., Fazio, D. and Carell, T. (2010) High-density DNA functionalization by a combination of Cu-catalyzed and Cu-free click chemistry. Chem. Eur. J. **16**, 6877–6883 CrossRef

53 Schoch, J., Staudt, M., Samanta, A., Wiessler, M. and Jäschke, A. (2012) Site-specific one-pot dual labeling of DNA by orthogonal cycloaddition chemistry. Bioconjug. Chem. **23**, 1382–1386 CrossRef PubMed

54 Schoch, J., Wiessler, M. and Jäschke, A. (2010) Post-synthetic modification of DNA by inverse-electron-demand Diels — Alder reaction. J. Am. Chem. Soc. **132**, 8846–8847 CrossRef PubMed

55 Gierlich, J., Burley, G.A., Gramlich, PME, Hammond, D.M. and Carell, T. (2006) Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. Org. Lett. **8**, 3639–3642 CrossRef PubMed

56 Marks, I.S., Kang, J.S., Jones, B.T., Landmark, K.J., Cleland, A.J. and Taton, T.A. (2011) Strain-promoted "Click" chemistry for terminal labeling of DNA. Bioconjug. Chem. **22**, 1259–1263 CrossRef PubMed

57 Crisalli, P., Hernández, A.R. and Kool, E.T. (2012) Fluorescence quenchers for hydrazone and oxime orthogonal bioconjugation. Bioconjug. Chem. **23**, 1969–1980 CrossRef PubMed

58 Gramlich, PME, Warncke, S, Gierlich, J. and Carell, T. (2008) Click–Click–Click: Single to triple modification of DNA. Angew. Chemie Int. Ed. **47**, 3442–3444 CrossRef

59 Burgess, K. and Cook, D. (2000) Syntheses of nucleoside triphosphates. Chem. Rev. **100**, 2047–2060 CrossRef PubMed

60 Ma, S., Saaem, I. and Tian, J. (2016) Error correction in gene synthesis technology. Trends Biotechnol **30**, 147–154 CrossRef