

# RSC Chemical Biology

rsc.li/rsc-chembio



ISSN 2633-0679

**COMMUNICATION**

Philipp Holliger, Alexander I. Taylor *et al.*  
Efficient synthesis and replication of diverse sequence  
libraries composed of biostable nucleic acid analogues

Cite this: *RSC Chem. Biol.*, 2022, 3, 1209Received 9th February 2022,  
Accepted 15th April 2022

DOI: 10.1039/d2cb00035k

rsc.li/rsc-chembio

# Efficient synthesis and replication of diverse sequence libraries composed of biostable nucleic acid analogues†

John R. D. Hervey,<sup>a</sup> Niklas Freund,<sup>b</sup> Gillian Houlihan,<sup>b</sup> Gurpreet Dhaliwal,<sup>a</sup> Philipp Holliger<sup>\*b</sup> and Alexander I. Taylor<sup>\*a</sup>

Functional nucleic acids can be evolved *in vitro* using cycles of selection and amplification, starting from diverse-sequence libraries, which are typically restricted to natural or partially-modified polymer chemistries. Here, we describe the efficient DNA-templated synthesis and reverse transcription of libraries entirely composed of serum nuclease resistant alternative nucleic acid chemistries validated in nucleic acid therapeutics; locked nucleic acid (LNA), 2'-O-methyl-RNA (2'OMe-RNA), or mixtures of the two. We evaluate yield and diversity of synthesised libraries and measure the aggregate error rate of a selection cycle. We find that in addition to pure 2'-O-methyl-RNA and LNA, several 2'OMe-RNA/LNA blends seem suitable and promising for discovery of biostable functional nucleic acids for biomedical applications.

## Introduction

Single-stranded nucleic acids can adopt sophisticated 3D structures, enabling functions as “chemical antibodies” (aptamers) and catalysts (ribo-, DNA- and XNAzymes), which form the basis of an array of biomedical technologies with potential applications in diagnosis and therapy.<sup>1–6</sup> As design of such functional sequences (beyond antisense reagents) *ab initio* is not yet possible, they must be discovered through *in vitro* selection and evolution, which typically involves cycles of synthesis, panning and recovery of sequence-diverse libraries (containing up to 10<sup>15</sup> variants), formalised as “systematic evolution ligands by exponential enrichment (SELEX)”.<sup>7,8</sup>

For applications *in vivo*, or in the presence of biological fluids, modified or non-natural nucleic acids (also known as

xeno nucleic acids, XNAs) are generally advantageous over DNA or RNA due to improved nuclease resistance inherent in alternative backbone chemistries with modified sugars and congeners.<sup>9–11</sup> Although post-SELEX modification of functional oligonucleotides with analogues is possible, this can reduce or abolish function and precludes the selection of novel XNA structures stable under physiological conditions. Instead, substitution of one or two of the four RNA or DNA nucleotides with analogues such as 2'-fluoro- and 2'-aminopyrimidines has been used to prepare partially-modified libraries for “modSELEX”,<sup>12,13</sup> yielding mixed-chemistry aptamers with improved biostability. However, such mixed chemistry aptamers (although generally more stable than DNA or RNA alone) remain vulnerable to nuclease degradation due to the presence of unmodified segments.

By comparison, comparatively few examples of “X-SELEX”<sup>14,15</sup> selections involving complete substitution of all four nucleoside triphosphates ((d)NTPs) with biostable artificial analogues (xNTPs) have been reported.<sup>16–25</sup> This has historically proven challenging due to the high substrate specificity of DNA and RNA polymerases, which – in most cases – must be evolved or engineered<sup>26–28</sup> to permit DNA-templated XNA polymerisation at full substitution for library synthesis. Although elegant DNA-tagging approaches have been devised to avoid the requirement for an XNA reverse transcriptase (RT),<sup>21,29,30</sup> where available it may be beneficial to utilise engineered RTs<sup>31</sup> for XNA-templated cDNA synthesis, enabling efficient amplification and preparation of templates for subsequent rounds of selection.

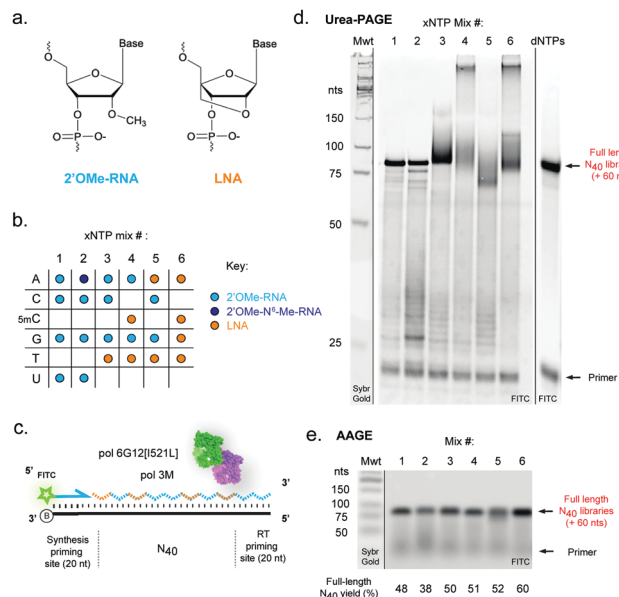
Among the different nucleic acid analogues, 2'-O-methyl-RNA (2'OMe-RNA), a natural post-transcriptional modification found in ribosomal, tRNA and mRNA, and 2'-O,4'-C-methylene-β-D-ribo- or ‘locked’ nucleic acid (LNA)<sup>32,33</sup> (Fig. 1a) are of particular interest. Both are resistant to serum nucleases and exhibit enhanced binding to complementary RNA and DNA, as well as formation of highly stable secondary structures, due to their conformationally restricted ribose ring structures (resulting from the 2' methoxy in 2'OMe-RNA and the bridging methylene group in LNA, positioned in the minor groove<sup>34,35</sup>),

<sup>a</sup> Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), University of Cambridge, Cambridge, CB2 0AW, UK. E-mail: ait29@cam.ac.uk

<sup>b</sup> Medical Research Council Laboratory of Molecular Biology, Cambridge, CB2 0QH, UK. E-mail: ph1@mrc-lmb.cam.ac.uk

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d2cb00035k>





**Fig. 1** Efficient templated synthesis of pure and mixed-chemistry diverse-sequence 2'OMe-RNA and LNA oligonucleotides. (a) Chemical structures and (b) nucleoside triphosphate (xNTP) combinations of 2'OMe-RNA and LNA used in the study. (c) Diagram and (d) denaturing acrylamide (PAGE) and (e) agarose (AAGE) gels showing DNA-templated synthesis of N<sub>40</sub> XNA libraries using the chemistries shown in (b), catalysed by a blend of engineered XNA polymerases “pol3M” and “pol6G12[I521L]”. Primer extensions were visualised by FITC fluorescence; unlabelled molecular weight standards (Mwt) were run on the same gels and revealed by staining with Sybr Gold. Gels are representative of at least three replicate reactions per chemistry.

which reduce the entropic penalty incurred by basepairing. Both 2'OMe-RNA and LNA have been shown to improve target strand invasion and specificity of antisense oligonucleotides, siRNAs and CRISPR/Cas systems, as well as boosting activity of DNazymes and aptamers in physiological conditions, and are generally well tolerated *in vivo*.<sup>11,36</sup> Furthermore, 2'OMe-RNA and LNA phosphoramidites are commercially available, enabling scalable chemical synthesis once functional sequences have been identified and optimised. Systems enabling the selection of biostable functional oligonucleotides using such chemistries therefore continue to be key technological goals for the field.

An early example of selections for aptamers composed of 2'OMe-RNA<sup>37</sup> where libraries were synthesised using a mutant T7 RNA polymerase (T7: Y639F, H784A)<sup>38</sup> required inclusion of unmodified GTP as well as ‘forcing’ conditions (high xNTP concentrations and Mn<sup>2+</sup>) in order to achieve synthesis of N<sub>30</sub> libraries. Furthermore, selection seemingly required this chimeric library to be supplemented with chemically-synthesised 2'OMe-RNA.<sup>37</sup> Reverse transcription of 2'OMe-RNA was possible with Thermoscript (an MMLV RT variant), although inefficient (~10% cDNA yields on 2'OMe-RNA templates), and the system had a high overall SELEX cycle error rate (51% of amplified cDNAs carried an error).<sup>37</sup> Subsequently, improvements in 2'OMe-RNA synthesis and RT were achieved using laboratory-evolved mutants of T7,<sup>39–41</sup> the Stoffel fragment of

Taq DNA polymerase<sup>20</sup> and KOD polymerase,<sup>25</sup> although these, too, required Mn<sup>2+</sup> and extended incubation times. Most recently, engineering a two-residue ‘steric gate’ in the Tgo DNA polymerase in order to reduce predicted clashes with bulky 2'-modified nucleotides yielded an efficient 2'OMe-RNA synthetase, “pol2M” and its variant “pol3M”.<sup>42</sup> Likewise, an efficient 2'OMe-RNA reverse transcriptase was engineered using Tgo by directed evolution, “RT-C8”.<sup>31</sup> In the case of LNA, a variety of polymerases have been explored for synthesis and RT,<sup>43</sup> including engineered variants of the polymerases Tgo (“polC7” and “RT521K”, respectively)<sup>17</sup> and KOD,<sup>25,44</sup> although LNA in selections has thus far been limited to primer regions<sup>45–47</sup> or one LNA nucleotide.<sup>25</sup>

Combinations of different XNAs in fully synthetic genetic systems (as demonstrated in ref. 19, 21, 25 and 48) offer a means to navigate a greater variety of chemical and structural space, potentially enabling discovery and evolution of more diverse functional 3D motifs and properties made possible by interactions between alternative polymer chemistries.<sup>49</sup> Here, we expand the range of mixed-chemistry synthetic genetic systems using two XNA polymerase blends to efficiently synthesise and reverse transcribe mixtures of LNA and/or 2'OMe-RNA.

Beyond sugar modifications, the addition of a variety of side chains to nucleobases, and the creation of alternative base-pairs, has proven to be a successful strategy for the expansion of chemical diversity and function of aptamer and DNazyme reagents.<sup>50–54</sup> However, their incorporation into selections has thus far been limited to DNA or RNA backbones. Combining base and sugar modifications<sup>55–57</sup> could conceivably offer a route to synthetic ligands and catalysts with improved function in biological contexts. We therefore also sought to explore the synthesis and reverse transcription of 2'OMe-RNA libraries also bearing a nucleobase modification, 2'-O-methyl-N<sup>6</sup>-methyl-A aka N<sup>6</sup>,2'-O-dimethyladenosine (m<sup>6</sup>A<sub>m</sub>), a natural terminal modification of eukaryotic mRNA involved in the regulation of transcript stability.<sup>58</sup> m<sup>6</sup>A<sub>m</sub> provides a methyl side-chain that could contribute to formation of hydrophobic paratopes in aptamers evolved to target proteins of interest by analogy to nature's use of m<sup>6</sup>A RNA modifications as sites for recognition by epitranscriptome regulatory proteins.<sup>59</sup>

Typically, synthesis and reverse transcription of XNAs is assayed by primer extension reactions with short, defined DNA templates and incorporation of a limited number of nucleotide analogues. However, to thoroughly evaluate the suitability of our systems for X-SELEX – in particular for more sophisticated directed evolution experiments such as aptamer selections against challenging targets – we examine synthesis and replication (*via* cDNA) of longer, diverse-sequence (N<sub>40</sub>) XNA libraries and assess yields, library diversity and replication fidelity following a complete X-SELEX cycle.

## Results and discussion

We reasoned that although reactions containing mixtures of 2'OMe-RNA and LNA nucleotides would in principle require



polymerase phenotypes capable of synthesis using both chemistries, a polymerase capable of efficiently incorporating one set of xNTPs could not necessarily be assumed to be capable of using the other with equal efficiency. Although it has been possible to identify a set of mutations that enable efficient synthesis of both chemistries in a single polymerase scaffold,<sup>25,42</sup> there are always trade-offs in efficiency and fidelity to be considered. Here, we have sought to explore a complementary approach to optimise synthesis using blends of XNA polymerases engineered in our labs. Blending polymerases is a well-established strategy for improving amplification of long or otherwise challenging DNA templates or in problematic reaction conditions.<sup>60–62</sup> Screening different combinations of XNA polymerases, we identified a two-polymerase blend that enabled efficient synthesis of mixtures of 2'OMe-RNA and LNA: pol3M<sup>42</sup> and another Tgo variant, “pol6G12[I521L]”,<sup>18</sup> previously developed for efficient synthesis of hexitol nucleic acid (HNA) in the absence of Mn<sup>2+</sup> (Tgo: V93Q, D141A, E143A, A485L, I521L, V589A, E609K, I610M, K659Q, E664Q, Q665P, R668K, D669Q, K671H, K674R, T676R, A681S, L704P, E730G). We initially screened all single- (each LNA-NTP into 2'OMe-NTP mixes) and double-nucleotide combinations (ESI,† Fig. S1), and identified a set of six xNTP mixes (Fig. 1b) that allowed efficient DNA-templated full-length synthesis of diverse-sequence (N<sub>40</sub>) polymers from 2'OMe-RNA primers (Fig. 1c–e) with good yields (expressed as % of yield with dNTPs (ESI,† Fig. S1)): all-2'OMe (75%); 2'OMe-m<sup>6</sup>A, 2'OMe-C, -G, -U (59%); 2'OMe-A, -C, -G, LNA-T (78%); 2'OMe-A, -G, LNA-5mC, -T (80%); LNA-A, -T, 2'OMe-C, -G (81%) and all-LNA (94%). Of the combinations explored, only an LNA purine/2'OMe-RNA pyrimidine mix was inefficiently synthesised (ESI,† Fig. S1).

As we and others have previously observed,<sup>17</sup> LNA × DNA template (as well as LNA × LNA inter- and intra-molecular) hybridisation can be only incompletely denatured by 8 M urea (and boiling in formamide loading buffer) during polyacrylamide gel electrophoresis (Urea-PAGE), resulting in low mobility species (Fig. 1d). We therefore also analysed samples by alkaline agarose gel electrophoresis (AAGE) and verified that synthesised libraries resolved into bands that were indeed *bona fide* full-length products (Fig. 1e) (subsequently confirmed by sequencing). As expected, all libraries were found to be highly biostable, with minimal degradation observed even after 5 days at 37 °C in 90% human serum (ESI,† Fig. S2), confirming the protection against enzymatic degradation afforded by 2'OMe-RNA and LNA is retained in the chimeric polymers.

To benchmark the pol3M/pol6G12[I521L] polymerase blend against alternative approaches for templated LNA and 2'OMe-RNA synthesis, we compared activity with the recently-described KOD polymerase variant “KOD DGLNK” specifically engineered for 2'OMe-RNA and LNA synthesis<sup>25</sup> (ESI,† Fig. S3). The blend consistently produced higher yields of pure and mixed-chemistry LNA and/or 2'OMe-RNA N<sub>40</sub> libraries than KOD DGLNK, in particular in the absence of Mn<sup>2+</sup> (ESI,† Fig. S3). We note that it remains possible that this performance advantage is at least partly due to our use of 2'OMe-RNA primers, which may be disfavoured by the KOD DGLNK variant.<sup>25</sup>

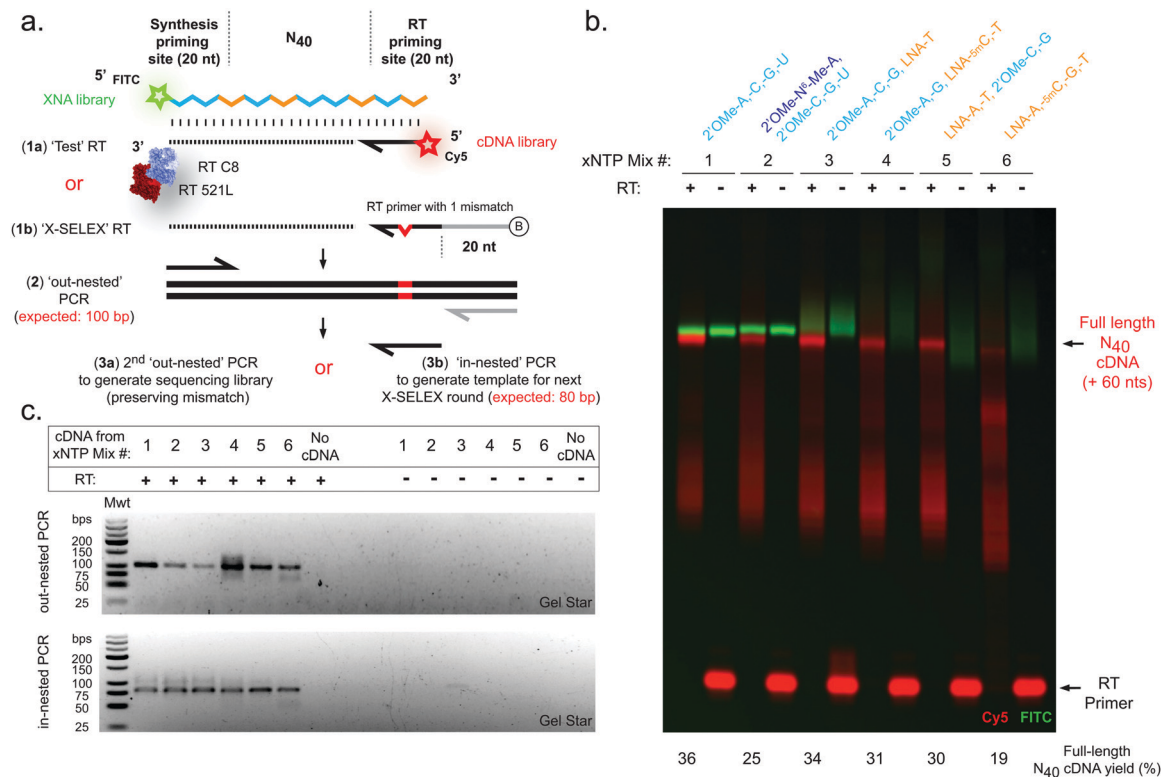
Next, we sought to explore reverse transcription of the 2'OMe-RNA/LNA libraries. A screen of XNA reverse transcriptases revealed that again a blend of two previously described RTs, RT-C8<sup>31</sup> and RT521L<sup>17</sup> (Tgo: V93Q, D141A, E143A, A385V, E429G, F445L, A485L, I521L, K726R), was capable of efficient synthesis of cDNA templated by N<sub>40</sub> libraries composed of all six 2'OMe-RNA and/or LNA combinations (Fig. 2), confirmed by both directly imaging cDNA (Fig. 2b) and a two-step semi-nested RT-PCR (Fig. 2a and c), the same amplification strategy used to generate templates for subsequent X-SELEX cycles.<sup>14,15</sup>

In contrast to efficient synthesis, reverse transcription from pure LNA N<sub>40</sub> templates (Fig. 2b) was clearly more challenging (19% cDNA yield) than pure 2'OMe-RNA (36% cDNA yield). This is not unexpected given that RT-C8 was specifically evolved for 2'OMe-RNA reverse transcription<sup>31</sup> and neither RT-C8 nor RT521L (identified by screening for activity on templates composed of HNA),<sup>17</sup> had been optimised for LNA RT. However, the mixed-chemistry libraries prepared using one or two LNA nucleoside triphosphates (with others 2'OMe-NTPs) were well tolerated (34%, 31% and 30% cDNA yield for the [LNA-T], [LNA-5mC, -T] and [LNA-A, -T] mixes, respectively) (Fig. 2b), as was the base-modified 2'OMe-N<sup>6</sup>A-containing mix (25% cDNA yield) (Fig. 2b). Specific amplicons could be easily obtained in PCRs templated by cDNA derived from all RT template chemistries (Fig. 2c). Although these yields suggest that XNA RT is the weakest link in the X-SELEX cycle, all systems nonetheless compare favourably with yields obtained with DNA-templated DNA synthesis assessed using Urea-PAGE (64%; ESI,† Fig. S1), and, as we have explored previously for the full 2'OMe-RNA system,<sup>31</sup> outperform previously engineered XNA RTs as well as commercially available polymerases.

Finally, we sought to confirm whether synthesis and reverse transcription reactions were indeed occurring in a templated manner with sufficient fidelity, and to assess whether sequence diversity (*i.e.* X-SELEX library quality) is maintained by the mixed-chemistry synthetic genetic systems. For these experiments, we chose to examine the complete replication cycle – *i.e.* the sum of synthesis, reverse transcription and cDNA amplification – rather than deconvolute the contributions of the individual polymerases as this aggregate measurement is a closer proxy of a full round of X-SELEX, and the performance of each polymerase has been described elsewhere.<sup>17,18,31,42</sup> An unbiased defined sequence (“Temp25”) (ESI,† Fig. S4), in addition to N<sub>40</sub> library sequences (Fig. 1 and 2), were therefore synthesised in each of the six 2'OMe-RNA and/or LNA systems, purified and reversed transcribed, and cDNA amplified to generate barcoded sequencing libraries for multiplexed deep sequencing (Fig. 3). RT reactions were primed using a DNA oligo with a single mismatch design (“XSELEX\_RT”; Fig. 2a and ESI,† Table S1, Fig. S4a), which ensured sequences analysed were derived from first-strand cDNA (excluding the possibility of contaminating DNA template from the XNA synthesis step).

Broadly, library diversity derived from the chemically-synthesised DNA template oligonucleotide (ESI,† Fig. S5) was efficiently maintained in all XNA systems (Fig. 3a), indicating a lack of any major biases in synthesis and reverse transcription.





**Fig. 2** Efficient reverse transcription of pure and mixed-chemistry diverse-sequence 2'OMe-RNA and LNA oligonucleotides. (a) Diagram showing strategies for assessment of XNA RT reactions. (b) Urea-PAGE gel showing XNA-templated (FITC, green) cDNA synthesis (Cy5, red). (c) Agarose gels showing amplification of cDNA by RT-PCR. Gels are representative of at least three replicate reactions per chemistry. RTs were catalysed by a blend of engineered XNA polymerases "RT-C8" and "RT521L" primed either by (a)(1a) a Cy5 fluorophore-labelled DNA primer ('Test' RT) in order to directly visualise cDNA synthesis (b), or alternatively RTs were primed by (a)(1b) a biotinylated DNA primer ('X-SELEX' RT) enabling purification of cDNA and amplification (c) via a two-step PCR strategy: firstly (a)(2) an 'out-nested' PCR dependent on a reverse primer site derived from the X-SELEX\_RT primer generates templates for, secondly, either (a)(3a) a PCR to generate sequencing libraries preserving a diagnostic mismatched base derived from the RT primer, or instead (a)(3b) an 'in-nested' PCR that regenerates the XNA synthesis template (although here is used only as a proof-of-concept demonstration).

In all systems, slight (~5%) overrepresentation of A was observed (although note that the template DNA N<sub>40</sub> oligo was found to have slightly higher A content as well (ESI,† Fig. S5)), except when the base-modified 2'OMe-N<sup>6</sup>A was used (Fig. 3a); a spike in As at the 40th nucleotide derives from single nucleotide deletions in either the template oligo (ESI,† Fig. S5) or during XNA synthesis, resulting in the 40th position being the first base of the conserved RT priming site (ESI,† Fig. S4). Conversely, in the pure LNA system, slight (5–10%) underrepresentation of G was observed. It is unclear if these variances are the result of differences in nucleotide analogue incorporation efficiency during synthesis, or are due to errors during reverse transcription, although the observation that nucleotide mixes in which 2'OMe-A was substituted for LNA-A, and/or 2'OMe-G for LNA-G were found to show generally lower synthesis yields than when 2'OMe-A and/or 2'OMe-G were used (ESI,† Fig. S1) would suggest that synthesis is the less efficient step, at least for LNA-G. If so, this may be compensated for by adjustment of the relative nucleotide analogue concentrations, although we did not explore this.

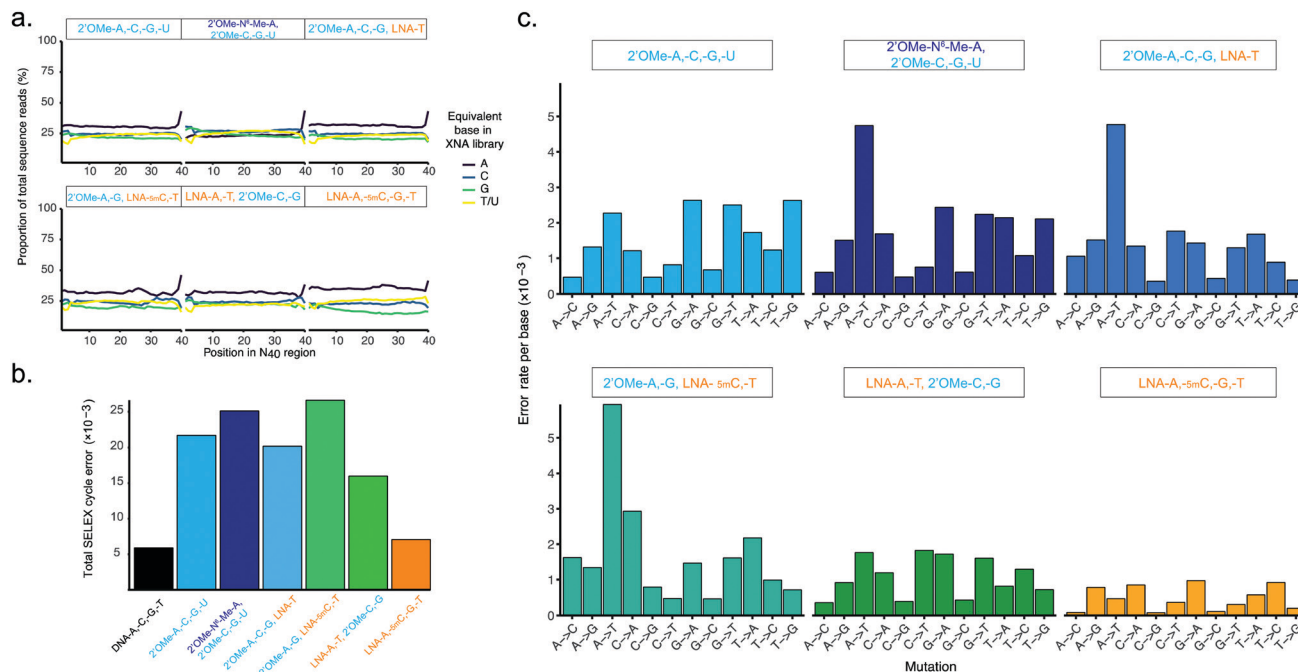
The fidelity of a complete cycle of synthesis and replication using the pol3M/pol6G12[1521L] and RT-C8/RT521L blends

(and a blend of Taq and the proof-reading polymerase Deep Vent for DNA amplification) was found to be similar between the pure 2'OMe-RNA ( $21.7 \times 10^{-3}$ ) and the mixed-chemistry systems ( $16.1\text{--}26.6 \times 10^{-3}$ ) (Fig. 3b and ESI,† Table S2), and generally comparable to (total X-SELEX cycle) error rates reported for analogous systems using KOD variants DGLNK and DLK<sup>25</sup> ( $15.6 \times 10^{-3}$  for pure 2'OMe-RNA,  $28.2 \times 10^{-3}$  for a mixed [2'OMe-A, -C, -G, LNA-T] system).

Interestingly, the mix containing both LNA-A and LNA-T, and the pure LNA system were found to have the lowest total error ( $16.1 \times 10^{-3}$  and  $7.05 \times 10^{-3}$ , respectively) (Fig. 3b and ESI,† Table S2). As this seemed remarkably low (comparable to a Tgo DNA-only system<sup>17</sup> ( $5.74 \times 10^{-3}$ ), and lower than the analogous KOD variant system<sup>25</sup> ( $14.0 \times 10^{-3}$ )), we also verified the fidelity of the pure LNA system by cloning and Sanger sequencing amplicons from the first step 'out-nest' PCR (rather than generating Illumina sequencing libraries) and obtained a similar total error ( $\sim 7.5 \times 10^{-3}$ ) (ESI,† Fig. S6).

The error profiles per base (Fig. 3c) suggest that, in the mixed-chemistry systems, misincorporation of As or Ts, and in particular A → T transversions, appear to be the dominant source of error. This is somewhat puzzling given the apparent





**Fig. 3** Pure and mixed-chemistry 2'OMe-RNA and LNA synthetic genetic systems enable synthesis and replication of biostable oligonucleotide libraries without substantial biases or loss of library diversity. Graphs show (a) diversity of  $N_{40}$  libraries (DNA shown in ESI† Fig. S5), and (b) aggregate error and (c) errors per base following a full cycle of synthesis of pure and mixed-chemistry 2'OMe-RNA and LNA oligonucleotides, reverse transcription and cDNA amplification (DNA → XNA → DNA → DNA sequenced). See also ESI† Tables S2 and S3 for further breakdown of errors and numbers of sequences analysed.

overrepresentation of As observed in the  $N_{40}$  sequences (Fig. 3a) and the apparent efficiency of synthesis of mixes containing 2'OMe-A (ESI† Fig. S1). However, a possible explanation could be the prevalence of AA dinucleotides (6 instances) and one AAA trinucleotide in the defined Temp25 sequence used for error analysis, which may be more challenging for 2'OMe-RNA systems than single incorporations. As these mixes also contain LNA-T, which appears to be easily incorporated (enabling the highest yields in single-LNA-nucleotide 2'OMe-RNA mixes (ESI† Fig. S1)), presumably pausing at the di- and trinucleotide positions then raises the incidence of LNA-T × dT mismatches (which appears not to be the case if 2'OMe-U is used instead, except in the system using 2'OMe-N<sup>6</sup>-Me-A, which is presumably more challenging than 2'OMe-A). Indeed, we find that the highest incidences of errors occur at AA(A) positions in the 2'OMe-RNA/LNA mixes (ESI† Table S3). Replacement of 2'OMe-A with LNA-A, as in the [LNA-A,-T, 2'OMe-C-G] mix, obviates this source of error and produces the highest fidelity mixed-chemistry system (Fig. 3b, c and ESI† Tables S2, S3).

In the pure LNA system, by contrast, A and T misincorporations are not a major source of errors (Fig. 3c). Likewise, we find no evidence of common (A → G/C → T) errors, which occur with the KOD variant LNA system presumably due to the increased stability of LNA × DNA wobble pairs (dT × LNA-G and/or dG × LNA-T)<sup>25</sup> (although we cannot rule out that these may be responsible for reduced efficiency of LNA-G incorporation during synthesis, resulting in the underrepresentation of G observed in the all-LNA library (Fig. 3a)). However, a commonly

overlooked caveat with such error analyses is the exclusion of early-terminated cDNA. Stalling during reverse transcription appears to occur more frequently with pure LNA templates than the mixed-chemistry or 2'OMe-RNA templates (Fig. 2b and ESI† Fig. S4c), resulting in cDNA that lacks the forward priming site and therefore fails to be amplified, sequenced and included in the analysis. It cannot therefore be ruled out that such errors do occur in our system, but do not propagate through the replication cycle. The error rates as measured nonetheless bode well for allowing enrichment of functional sequences in X-SELEX experiments, provided selection steps yield sufficient XNA templates to overcome the reduced reverse transcription efficiency; indeed, as we report elsewhere, functional biostable oligonucleotides ('2'OMezyme' catalysts) could be readily evolved from an all-2'OMe-RNA system (using pol3M or a functionally similar variant pol2M for synthesis, and RT-C8 for reverse transcription).<sup>42</sup>

## Conclusions

Using blends of previously engineered XNA polymerases, we have established a series of synthetic genetic systems based on mixtures of 2'OMe-RNA and LNA. We report the efficient synthesis and (indirect) replication of diverse-sequence oligonucleotide analogue libraries without major biases in incorporation or reverse transcription. All systems show similar efficiency and fidelity as an all-2'OMe-RNA system, which we



have already successfully used to select 2'OMezymes,<sup>42</sup> and our data suggest that systems in which 2'OMe-NTPs are substituted by one (LNA-T), two (LNA-A,-T) or all four LNA-NTPs offer further improvements. We provide a key proof-of-concept of a system bearing both sugar and nucleobase modifications, suggesting that selections for biostable functional oligos with expanded base chemistry are feasible using commercially available nucleotides (2'OMe-N<sup>6</sup>-methyl-ATP, 2'OMe-CTP, -GTP and -UTP), paving the way for development of XNA selection systems with more elaborate side chains.

The establishment of a variety of XNA systems based on highly biostable chemistries offers promising platform technologies for the development of a wide range of biostable functional oligonucleotides for diagnostic and therapeutic applications *in vivo*.<sup>63,64</sup> The use of oligo libraries fully resistant to serum nucleases minimises or abolishes the need to apply extensive post-selection modifications to prepare resulting functional oligonucleotides for *in vivo* applications (which can negatively affect activity), and enhances the prospects for selections in more realistic biological settings: in live cell or organoid culture or even in whole animals.<sup>65</sup> Excitingly, such approaches may offer the prospect of selections for aptamers capable of cell- or organ-specific delivery, and could enable direct screening for modulators of biological phenotype, a crucial approach in the immunoglobulin antibody discovery pipeline,<sup>66</sup> but which has yet to be implemented for oligonucleotide aptamers and catalysts.

## Author contributions

A. I. T. conceived and directed the project. J. R. D. H. acquired, curated and analysed data with A. I. T. and G. D. Polymerases were supplied by N. F., G. H. and P. H. A. I. T. and J. R. D. H. wrote the manuscript with contributions from the other authors.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

The authors thank Prof. J. Wengel (SDU, Denmark) for providing LNA-NTPs and Dr B. Porebski (MRC LMB, Cambridge, UK) for advice on sequence analysis. This work was supported by the Wellcome Trust and Royal Society through a Sir Henry Dale Fellowship to A. I. T (215453/Z/19/Z), the Boehringer Ingelheim Fonds through a PhD fellowship to N. F., the Medical Research Council (MRC) through a program grant to P. H. (MC\_U105178804) and the NIHR Cambridge BRC through a pump priming award to A. I. T.

## References

- 1 L. Ma and J. Liu, *iScience*, 2020, **23**, 100815.
- 2 R. Micura and C. Hobartner, *Chem. Soc. Rev.*, 2020, **49**, 7331–7353.
- 3 S. M. Nimjee and B. A. Sullenger, *Curr. Med. Chem.*, 2020, **27**, 4181–4193.
- 4 L. Li, S. Xu, H. Yan, X. Li, H. S. Yazd, X. Li, T. Huang, C. Cui, J. Jiang and W. Tan, *Angew. Chem., Int. Ed.*, 2021, **60**, 2221–2231.
- 5 H. Peng, B. Latifi, S. Muller, A. Luptak and I. A. Chen, *RSC Chem. Biol.*, 2021, **2**, 1370–1383.
- 6 S. Shigdar, B. Schrand, P. H. Giangrande and V. de Francis, *Mol. Ther.*, 2021, **29**, 2396–2411.
- 7 A. D. Ellington and J. W. Szostak, *Nature*, 1990, **346**, 818–822.
- 8 C. Tuerk and L. Gold, *Science*, 1990, **249**, 505–510.
- 9 A. I. Taylor, G. Houlihan and P. Holliger, *Cold Spring Harbor Perspect. Biol.*, 2019, **11**, a032490.
- 10 K. Duffy, S. Arangundy-Franklin and P. Holliger, *BMC Biol.*, 2020, **18**, 112.
- 11 L. K. McKenzie, R. El-Khoury, J. D. Thorpe, M. J. Damha and M. Hollenstein, *Chem. Soc. Rev.*, 2021, **50**, 5126–5164.
- 12 Y. Lin, Q. Qiu, S. C. Gill and S. D. Jayasena, *Nucleic Acids Res.*, 1994, **22**, 5229–5234.
- 13 A. D. Keefe and S. T. Cload, *Curr. Opin. Chem. Biol.*, 2008, **12**, 448–456.
- 14 A. I. Taylor and P. Holliger, *Curr. Protoc. Chem. Biol.*, 2018, **10**, e44.
- 15 A. I. Taylor and P. Holliger, *Nat. Protoc.*, 2015, **10**, 1625–1642.
- 16 S. Jhaveri, B. Olwin and A. D. Ellington, *Bioorg. Med. Chem. Lett.*, 1998, **8**, 2285–2290.
- 17 V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S.-Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn and P. Holliger, *Science*, 2012, **336**, 341–344.
- 18 A. I. Taylor, V. B. Pinheiro, M. J. Smola, A. S. Morgunov, S. Peak-Chew, C. Cozens, K. M. Weeks, P. Herdewijn and P. Holliger, *Nature*, 2015, **518**, 427–430.
- 19 A. D. Friedman, D. Kim and R. Liu, *Biomaterials*, 2015, **36**, 110–123.
- 20 Z. Liu, T. Chen and F. E. Romesberg, *Chem. Sci.*, 2017, **8**, 8179–8182.
- 21 S. Arangundy-Franklin, A. I. Taylor, B. T. Porebski, V. Genna, S. Peak-Chew, A. Vaisman, R. Woodgate, M. Orozco and P. Holliger, *Nat. Chem.*, 2019, **11**, 533–542.
- 22 E. Ereemeeva, A. Fikatas, L. Margamuljana, M. Abramov, D. Schols, E. Groaz and P. Herdewijn, *Nucleic Acids Res.*, 2019, **47**, 4927–4939.
- 23 M. R. Dunn, C. M. McCloskey, P. Buckley, K. Rhea and J. C. Chaput, *J. Am. Chem. Soc.*, 2020, **142**, 7721–7724.
- 24 I. Alves Ferreira-Bravo and J. J. DeStefano, *Viruses*, 2021, **13**, 1983–1995.
- 25 H. Hoshino, Y. Kasahara, M. Kuwahara and S. Obika, *J. Am. Chem. Soc.*, 2020, **142**, 21530–21537.
- 26 G. Houlihan, S. Arangundy-Franklin and P. Holliger, *Curr. Opin. Biotechnol.*, 2017, **48**, 168–179.
- 27 G. Houlihan, S. Arangundy-Franklin and P. Holliger, *Acc. Chem. Res.*, 2017, **50**, 1079–1087.



- 28 N. Freund, M. Furst and P. Holliger, *Curr. Opin. Biotechnol.*, 2021, **74**, 129–136.
- 29 J. K. Ichida, K. Zou, A. Horhota, B. Yu, L. W. McLaughlin and J. W. Szostak, *J. Am. Chem. Soc.*, 2005, **127**, 2802–2803.
- 30 H. Yu, S. Zhang and J. C. Chaput, *Nat. Chem.*, 2012, **4**, 183–187.
- 31 G. Houlihan, S. Arangundy-Franklin, B. T. Porebski, N. Subramanian, A. I. Taylor and P. Holliger, *Nat. Chem.*, 2020, **12**, 683–690.
- 32 S. Obika, D. Nanbu, Y. Hari, K. Morio, Y. In, T. Ishida and T. Imanishi, *Tetrahedron Lett.*, 1997, **38**, 8735–8738.
- 33 A. A. Koshkin, S. K. Singh, P. Nielsen, V. K. Rajwanshi, R. Kumar, M. Meldgaard, C. E. Olsen and J. Wengel, *Tetrahedron Lett.*, 1998, **54**, 3607–3630.
- 34 K. E. Nielsen, S. K. Singh, J. Wengel and J. P. Jacobsen, *Bioconjugate Chem.*, 2000, **11**, 228–238.
- 35 M. Petersen, K. Bondensgaard, J. Wengel and J. P. Jacobsen, *J. Am. Chem. Soc.*, 2002, **124**, 5974–5982.
- 36 S. T. Crooke, X. H. Liang, B. F. Baker and R. M. Crooke, *J. Biol. Chem.*, 2021, **296**, 100416.
- 37 P. E. Burmeister, S. D. Lewis, R. F. Silva, J. R. Preiss, L. R. Horwitz, P. S. Pendergrast, T. G. McCauley, J. C. Kurz, D. M. Epstein, C. Wilson and A. D. Keefe, *Chem. Biol.*, 2005, **12**, 25–33.
- 38 R. Padilla and R. Sousa, *Nucleic Acids Res.*, 2002, **30**, e138.
- 39 J. Ibach, L. Dietrich, K. R. M. Koopmans, N. Nöbel, M. Skoupi and S. Brakmann, *J. Biotechnol.*, 2013, **167**, 287–295.
- 40 A. J. Meyer, D. J. Garry, B. Hall, M. M. Byrom, H. G. McDonald, X. Yang, Y. W. Yin and A. D. Ellington, *Nucleic Acids Res.*, 2015, **43**, 7480–7488.
- 41 J. W. Ellefson, J. Gollihar, R. Shroff, H. Shivram, V. R. Iyer and A. D. Ellington, *Science*, 2016, **352**, 1590–1593.
- 42 N. Freund, A. I. Taylor, S. Arangundy-Franklin, N. Subramanian, S.-Y. Peak-Chew, A. M. Whitaker, B. D. Freudenthal, M. Abramov, P. Herdewijn and P. Holliger, *Nat. Chem.*, 2022, in press.
- 43 K. K. Karlsen and J. Wengel, *Nucleic Acid Ther.*, 2012, **22**, 366–370.
- 44 R. N. Veedu, B. Vester and J. Wengel, *Org. Biomol. Chem.*, 2009, **7**, 1404–1409.
- 45 I. C. Elle, K. K. Karlsen, M. G. Terp, N. Larsen, R. Nielsen, N. Derbyshire, S. Mandrup, H. J. Ditzel and J. Wengel, *Mol. Biosyst.*, 2015, **5**, 1260–1270.
- 46 Y. Kasahara, Y. Irisawa, H. Fujita, A. Yahara, H. Ozaki, S. Obika and M. Kuwahara, *Anal. Chem.*, 2013, **85**, 4961–4967.
- 47 M. Kuwahara and S. Obika, *Artificial DNA: PNA & XNA*, 2013, **4**, 39–48.
- 48 D. Thirunavukarasu, T. Chen, Z. Liu, N. Hongdilokkul and F. E. Romesberg, *J. Am. Chem. Soc.*, 2017, **139**, 2892–2895.
- 49 T. Efthymiou, J. Gavette, M. Stoop, F. De Riccardis, M. Froeyen, P. Herdewijn and R. Krishnamurthy, *Chemistry*, 2018, **24**, 12811–12819.
- 50 J. C. Rohloff, A. D. Gelinas, T. C. Jarvis, U. A. Ochsner, D. J. Schneider, L. Gold and N. Janjic, *Mol. Ther.–Nucleic Acids*, 2014, **3**, e201.
- 51 F. Pfeiffer, M. Rosenthal, J. Siegl, J. Ewers and G. Mayer, *Curr. Opin. Biotechnol.*, 2017, **48**, 111–118.
- 52 Y. Wang, E. Liu, C. H. Lam and D. M. Perrin, *Chem. Sci.*, 2018, **9**, 1813–1821.
- 53 E. Biondi and S. A. Benner, *Biomedicines*, 2018, **6**, 53–65.
- 54 Y. W. Cheung, P. Rothlisberger, A. E. Mechaly, P. Weber, F. Levi-Acobas, Y. Lo, A. W. C. Wong, A. B. Kinghorn, A. Haouz, G. P. Savage, M. Hollenstein and J. A. Tanner, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 16790–16798.
- 55 H. Mei and J. C. Chaput, *Chem. Commun.*, 2018, **54**, 1237–1240.
- 56 M. Renders, S. Dumbre, M. Abramov, D. Kestemont, L. Margamuljana, E. Largy, C. Cozens, J. Vandenameele, V. B. Pinheiro, D. Toye, J. M. Frere and P. Herdewijn, *Nucleic Acids Res.*, 2019, **47**, 2160–2168.
- 57 F. Levi-Acobas, A. Katolik, P. Rothlisberger, T. Cokelaer, I. Sarac, M. J. Damha, C. J. Leumann and M. Hollenstein, *Org. Biomol. Chem.*, 2019, **17**, 8083–8087.
- 58 J. Mauer, X. Luo, A. Blanjoie, X. Jiao, A. V. Grozhik, D. P. Patil, B. Linder, B. F. Pickering, J. J. Vasseur, Q. Chen, S. S. Gross, O. Elemento, F. Debart, M. Kiledjian and S. R. Jaffrey, *Nature*, 2017, **541**, 371–375.
- 59 D. P. Patil, B. F. Pickering and S. R. Jaffrey, *Trends Cell Biol.*, 2018, **28**, 113–127.
- 60 W. M. Barnes, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 2216–2220.
- 61 J. Hedman, A. Nordgaard, C. Dufva, B. Rasmusson, R. Ansell and P. Radstrom, *Anal. Biochem.*, 2010, **405**, 192–200.
- 62 D. Millar, Y. Christova and P. Holliger, *Nucleic Acids Res.*, 2015, **43**, e155.
- 63 J. Zhou and J. Rossi, *Nat. Rev. Drug Discovery*, 2017, **16**, 181–202.
- 64 A. A. Fokina, D. A. Stetsenko and J.-C. François, *Expert Opin. Biol. Ther.*, 2015, **15**, 689–711.
- 65 M. Sola, A. P. Menon, B. Moreno, D. Meraviglia-Crivelli, M. M. Soldevilla, F. Carton-Garcia and F. Pastor, *Mol. Ther.–Nucleic Acids*, 2020, **21**, 192–204.
- 66 R. R. Minter, A. M. Sandercock and S. J. Rust, *Drug Discovery Today Technol.*, 2017, **23**, 83–90.

