

Cite this: *Chem. Sci.*, 2025, 16, 1966

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 13th October 2024  
Accepted 17th December 2024

DOI: 10.1039/d4sc06958g

rsc.li/chemical-science

# Long oligos: direct chemical synthesis of genes with up to 1728 nucleotides†

Yipeng Yin,<sup>‡a</sup> Reed Arneson,<sup>‡b</sup> Yinan Yuan<sup>\*b</sup> and Shiyue Fang<sup>ID</sup> <sup>\*a</sup>

The longest oligos that can be chemically synthesized are considered to be 200-mers. Here, we report direct synthesis of an 800-mer green fluorescent protein gene and a 1728-mer  $\phi$ 29 DNA polymerase gene on an automated synthesizer. Key innovations that enabled this breakthrough include conducting the synthesis on a smooth surface rather than within the pores of traditional supports, and the use of the powerful catching-by-polymerization (CBP) method for isolating the full-length oligos from a complex mixture. Conducting synthesis on a smooth surface not only eliminated the steric hindrance that would otherwise prevent long oligo assembly, but also, surprisingly, drastically reduced synthesis errors. Compared with the benchmark PCR assembly gene synthesis method, the direct long oligo synthesis method has the advantages of higher probability to succeed, fewer sequence restrictions, and being able to synthesize long oligos containing difficult elements such as unusually stable higher-order structures, long repeats, and site-specific modifications. The method is expected to open doors for various projects in areas such as synthetic biology, gene editing, and protein engineering.

## Introduction

Many areas such as synthetic biology,<sup>1</sup> nucleic acid vaccines,<sup>2</sup> oligonucleotide therapeutics,<sup>3</sup> CRISPR Cas9 gene editing,<sup>4</sup> and protein engineering,<sup>5</sup> require *de novo* synthesis of DNAs. In many of the cases, the DNAs need to be longer than 200 nucleotides (nt).<sup>6</sup> Because the state of the art chemical synthesis methods cannot reliably produce oligos longer than 200 nt,<sup>7</sup> the DNAs have to be produced *via* biological means such as PCR assembly or less commonly, ligation using synthetic oligos shorter than 200-mers.<sup>8</sup> While biological methods have provided the required DNAs for the areas to emerge and advance, many sequences desperately needed by some projects are beyond the reach of existing methods.<sup>9</sup> For example, if a sequence contains higher order structures with unusual stability, the PCR assembly method may not function effectively. If a sequence contains repeating segments that are longer than the short oligos available from chemical synthesis, biological means would not be able to accurately assemble the target DNA. If a sequence contains one or more site-specific modifications such as m<sup>6</sup>A, the PCR assembly method may not

be able to produce the sequence. To overcome these and other challenges, which were discussed in a previous article,<sup>10</sup> there is a need to develop new methods for direct *de novo* synthesis of oligos longer than 200-mers.

The most notable achievement in the area of *de novo* long oligo synthesis in recent years is the development of the template-independent enzymatic oligo synthesis (TiEOS) technologies, primarily utilizing engineered terminal deoxy-nucleotidyl transferases (TdT).<sup>7,11–15</sup> While these technologies hold great promise, they are not without shortcomings. For example, the large enzyme-to-nucleotide mass ratio is not atomically economic, which may be one of the reasons for the high cost of the methods if the enzyme is not used in a catalytic quantity or recycled. The higher-order structures of long oligos may reduce synthesis efficiency.<sup>12</sup> The coupling time may be lengthy, and the coupling yield may not meet the expectations for typical enzymatic reactions.<sup>12</sup> In addition, enzymatic methods typically lack a capping step, increasing the likelihood of deletion errors.<sup>13</sup> The TdT enzyme exhibits inherent nucleotide biases, leading to lower coupling efficiency for certain nucleotides, a problem that may be difficult to overcome through enzyme engineering.<sup>12</sup> Finally, the method may be difficult to be adapted for synthesizing long oligos with site-specific modifications.

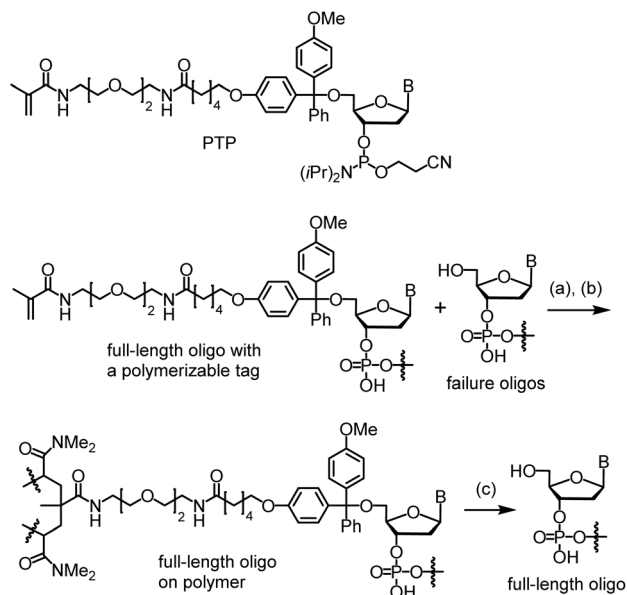
In contrast to the resources invested in developing enzymatic methods for long oligo synthesis, little effort has been dedicated to advancing chemical methods over the past decade, even though many of the aforementioned shortcomings of enzymatic methods may be addressable using chemical approaches. Since 2010, our research team has been making efforts to develop

<sup>a</sup>Department of Chemistry, , and Health Research Institute, Michigan Technological University, Houghton, Michigan, 49931, USA. E-mail: shifang@mtu.edu

<sup>b</sup>College of Forest Resources and Environmental Science, Michigan Technological University, Houghton, Michigan, 49931, USA. E-mail: yinyuan@mtu.edu

† Electronic supplementary information (ESI) available: Experimental details, glass wool and glass bead loading calculation, and sequencing results for oligos synthesized on glass wool, glass beads and CPG. See DOI: <https://doi.org/10.1039/d4sc06958g>

‡ Equal contributors.



**Scheme 1** Catching-by-polymerization (CBP). Conditions: (a) *N,N*-dimethylacrylamide, *N,N'*-methylenebis(acrylamide), sodium acrylate, ammonium persulfate, *N,N,N',N'*-tetramethylethylenediamine (TMEDA), rt, 1 h. (b) Wash. (c) AcOH (80%), rt, 5 min  $\times$  3. PTP = polymerizable tagging phosphoramidite. B = nucleobase.

a method called catching-by-polymerization (CBP) for synthetic oligo purification (Scheme 1).<sup>16</sup> The method involves tagging the full-length oligo with a polymerizable tagging phosphoramidite (PTP) and incorporating the conjugate into a polyacrylamide gel. Because failure oligos are capped during automated synthesis, they are not tagged, and therefore are not incorporated into the gel. Oligo purification can thus be achieved by washing away the failure oligos, followed by cleaving the full-length oligo from the gel. Recognizing the power of the CBP method, we attempted to use it to isolate the extremely low percentage (but sufficient quantities) of full-length oligos from the complex mixture generated from the thousands of reactions required for long oligo synthesis.<sup>17,18</sup> Most recently, using the

CBP method, we succeeded in purification of 400-mers. Sanger sequencing confirmed the sequences.<sup>10</sup> Here, with additional innovations involving the use of glass wool and glass beads as a solid support for long oligo synthesis, we report direct chemical synthesis of the 800-mer green fluorescent protein (GFP) gene and the 1728-mer  $\Phi$ 29 DNA polymerase gene, and their isolation with CBP and characterization with Sanger sequencing.

## Results

Previously, we successfully synthesized 400-mer oligos and employed the CBP method to isolate the full-length sequences from the complex mixture of crude products.<sup>10,19</sup> The key elements that contributed to the success include using controlled pore glass (CPG) with pore sizes as large as 2000 Å as the solid support and reducing its loading by inactivating a portion of the reactive sites. These adjustments reduced the steric hindrance of the CPG, which we believe is crucial for the efficient synthesis of long oligos. Realizing the unlimited power of the CBP method for isolating low percentages of full-length oligos in the course of that work, we reasoned that if the steric hindrance can be further reduced, even longer oligos can be synthesized and isolated. Instead of reducing the loading of CPG further, we considered conducting the synthesis on a smooth surface as opposed to within pores of solid supports. A typical concern for this approach is insufficient loading, which is the reason for traditional oligo synthesis to be conducted in pores.

With the potential loading problem in mind, we thought that glass wool would partially solve the problem. Therefore, we calculated the loading of glass wool, and made a comparison with that of glass beads (Table 1).

Glass wool with a diameter of  $\sim 8 \mu\text{m}$  is commercially available and inexpensive. We tested its resistance to liquid flow, and found that it is virtually resistance-free, which is required for solid phase synthesis. Assuming a length of 1 cm, density of  $2.2 \text{ g ml}^{-1}$ , and  $3.2 \text{ molecules per nm}^2$  (Table 1),<sup>22</sup> the loading is

**Table 1** Loading of solid supports, and yields of long oligo synthesis

| Entry | Items                             | Glass wool <sup>a</sup>                       | Glass beads                                   | CPG 2K Å tested               | CPG 2K Å                              | Wang resin                             |
|-------|-----------------------------------|-----------------------------------------------|-----------------------------------------------|-------------------------------|---------------------------------------|----------------------------------------|
| 1     | Density <sup>b</sup>              | $2.2 \text{ g ml}^{-1}$                       | $2.2 \text{ g ml}^{-1}$                       |                               |                                       |                                        |
| 2     | Diameter <sup>c</sup>             | $8 \mu\text{m}$                               | $58 \mu\text{m}$                              |                               |                                       |                                        |
| 3     | Loading formula <sup>d</sup>      | $10.6 \div (d \times r) \mu\text{mol g}^{-1}$ | $15.9 \div (d \times r) \mu\text{mol g}^{-1}$ |                               |                                       |                                        |
| 4     | Loading                           | $1.208 \mu\text{mol g}^{-1e}$                 | $0.249 \mu\text{mol g}^{-1e}$                 | $5.405 \mu\text{mol g}^{-1f}$ | $20\text{--}30 \mu\text{mol g}^{-1g}$ | $0.3\text{--}2.5 \text{ mmol g}^{-1g}$ |
| 5     | Relative loading                  | 4.8                                           | 1                                             | 22                            | $\sim 100^h$                          | $\sim 10\,000^i$                       |
| 6     | Measured loading <sup>j</sup>     | $0.981 \mu\text{mol g}^{-1}$                  | $0.256 \mu\text{mol g}^{-1}$                  | $5.359 \mu\text{mol g}^{-1}$  |                                       |                                        |
| 7     | 800-mer synthesized <sup>k</sup>  | $3.7 \text{ nmol g}^{-1}$                     | $0.034 \text{ nmol g}^{-1}$                   |                               |                                       |                                        |
| 8     | 800-mer yield                     | $3.7/981 = 0.38\%$                            | $0.034/256 = 0.013\%$                         |                               |                                       |                                        |
| 9     | 1728-mer synthesized <sup>k</sup> |                                               | $0.041 \text{ nmol g}^{-1}$                   |                               |                                       |                                        |
| 10    | 1728-mer yield                    |                                               | $0.041/256 = 0.016\%$                         |                               |                                       |                                        |

<sup>a</sup> The length of glass wool can be 1 cm or longer. <sup>b</sup> Density is that of solid glass. <sup>c</sup> To achieve close to resistance-free flow of liquid, the diameter of glass beads needs to be  $\sim 50 \mu\text{m}$  or larger. <sup>d</sup> The effect of the length of glass wool on loading is minimal and is omitted in the formula. The units for  $d$  and  $r$  in the formulae are  $\text{g ml}^{-1}$  and  $\mu\text{m}$ , respectively. Details for deriving the formulae are in the ESI. <sup>e</sup> Calculated value assuming  $3.2 \text{ molecules per nm}^2$ . <sup>f</sup> Given by the manufacturer of the CPG tested. <sup>g</sup> Values from the literature.<sup>20,21</sup> <sup>h</sup>  $25 \mu\text{mol g}^{-1}$  is used for the calculation. <sup>i</sup>  $2.5 \text{ mmol g}^{-1}$  is used for the calculation. <sup>j</sup> Measure using trityl assay using the literature procedure.<sup>20</sup> <sup>k</sup> Oligo obtained per gram of solid support after CBP purification as determined with a Qubit 4 Fluorometer.



1208 nmol g<sup>-1</sup> (see the ESI† for calculations). For glass beads, to allow for close to resistance-free liquid flow, ideally their diameter is ~50 μm or larger. Assuming a diameter of 58 μm and a density of 2.2 g ml<sup>-1</sup>, the loading is calculated to be 249 nmol g<sup>-1</sup>. Therefore, the loading of glass wool is ~4.8 times that of glass beads (entry 5).

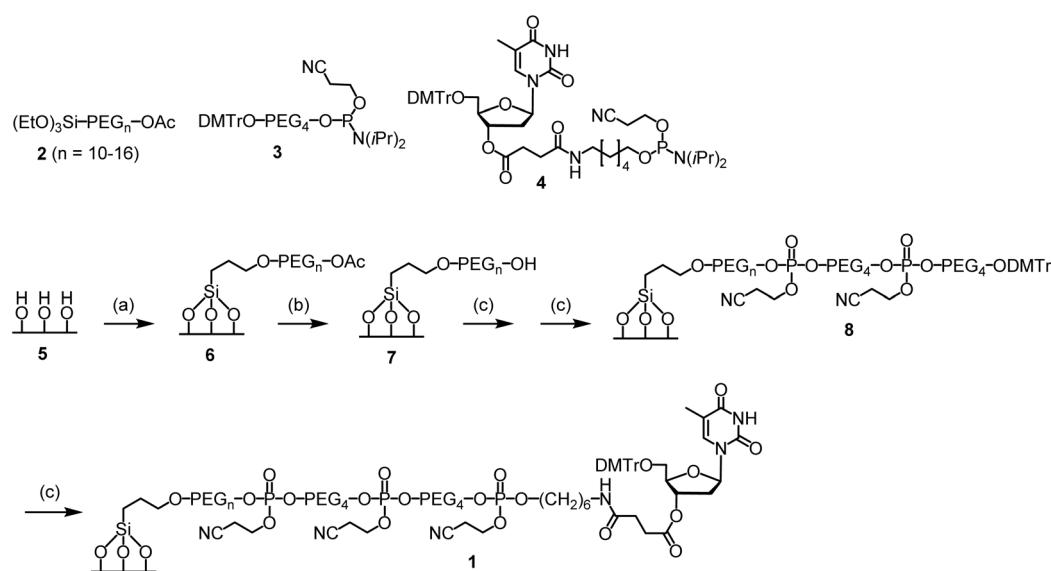
For comparison, the loading of commercial CPG with 2000 Å diameter is typically 20–30 μmol g<sup>-1</sup>,<sup>20</sup> which is ~100 times higher than that of glass beads (entry 5). The loading of the Wang resin (widely used for peptide synthesis) is 0.3–2.5 mmol g<sup>-1</sup>,<sup>21</sup> which is close to 10 000 times higher. However, for long oligo synthesis, we reasoned that low loading is less of an issue. For most biological applications, as little as 1 pmol oligo is sufficient.<sup>23,24</sup> Using glass wool, with a 100 mg support, which is the quantity that can be directly used under typical small scale oligo synthesis conditions, assuming an average stepwise yield of 99.7%, which corresponds to an overall yield of 0.25% for a 2000-mer synthesis, the quantity of full-length oligo is ~296 pmol, which is much larger than 1 pmol. However, the low percentage yield is a serious problem because there is no method to purify or concentrate the full-length oligo. For example, HPLC would not be able to resolve the full-length oligo from failure ones. Gel electrophoresis would not be able to resolve this either and even if it could be engineered to resolve, the full-length oligo would be invisible on the gel due to its low percentage. Solid phase extraction methods<sup>25–28</sup> may not be suitable for the task either because the high entropy barrier for reactions between large molecules and reactive sites on a solid surface would make the extraction inefficient, and it may be difficult for the large molecules to enter the pores of the solid support in the first place. However, using CBP, the low percentage problem can be overcome. With these considerations, we went ahead and synthesized long oligos on glass wool using the 800-mer GFP gene as an example.

## Glass wool functionalization

To conduct the synthesis, the functionalized glass wool **1** was prepared (Scheme 2). The required compounds **2** and **3** were purchased from commercial sources. Compound **4** was synthesized on-site (ESI and Scheme S1†). Glass wool (**5**) was activated by treating with a piranha solution under previously used conditions.<sup>29</sup> The surface was then PEGylated using **2** to give **6**, the acetyl group of which was removed with ammonia to give **7**.<sup>30</sup> For solid phase synthesis, when the reaction site is close to the surface, the reaction is less efficient. Therefore, the linker in **7** was elongated using compound **3**. The elongation reactions were carried out on a MerMade 6 synthesizer under typical DNA synthesis conditions but with a longer reaction time. After two cycles, **8** was obtained, which was converted to the target functionalized glass wool **1** under similar conditions on the synthesizer using compound **4** as the phosphoramidite monomer. More details regarding the synthesis are provided in the ESI† The success of the surface functionalization reactions could be estimated timely by observing the color of trityl cations during the detritylation steps (Fig. S1†). The loading of the glass wool thus functionalized was determined to be 981 nmol g<sup>-1</sup> (entry 6, Table 1) using a reported method,<sup>20</sup> which is not far away from the calculated value of 1208 nmol g<sup>-1</sup> (entry 4).

## Synthesis of an 800-mer oligo on glass wool

The synthesis of the 800-mer GFP gene on glass wool was carried out under typical conditions of phosphoramidite chemistry with little modification. The functionalized glass wool (30 mg, 29.4 nmol) was packed into an empty 0.2 μmol synthesis column. Even though the scale of the synthesis was only 29.4 nmol, to ensure complete coverage of the glass wool in every reaction in the entire synthesis, the synthesizer manufacturer recommended 1 μmol synthetic cycle was used. While the conditions



**Scheme 2** Functionalization of glass wool and glass beads. Conditions: (a) **2** (1% PhMe), rt, 20 min; then, supernatant removed, glass wool 100 °C, 4 h. (b) NH<sub>4</sub>OH (30%), 55 °C, 2 h. (c) On a DNA synthesizer, standard coupling, oxidation and deblocking conditions with modifications; see the ESI† for details.



may be engineered to minimize reagent use considering that less reagent is needed to fill the synthesis column for the case of glass wool, which does not need a reagent to fill pores, than that of traditional porous supports, that engineering studies was not pursued in the present study. The synthesis was tested on both ABI-394 and MerMade 6 synthesizers. Even though the former consumed more solvents, it was preferred due to its shorter synthesis time. In principle, the synthesis can be carried out consecutively, which would only need a little more than three days, we had to pause the synthesis two or more times to refill reagents. With the pauses, the synthesis was completed within four days. Trityl assay consistently gave 99.6% to 99.8% average stepwise yields once the synthetic cycle ran over 100 times consecutively. To tag the full-length oligo for CBP purification, the last nucleotide at the 5'-end of the 800-mer was not included in the above synthesis procedure. Instead, the 799-mer with a trityl group at its 5'-end on glass wool was transferred to a column suitable for a MerMade 6 synthesizer. Upon delivery of the deblocking agent, the orange color characteristic of dilute trityl cations could still be observed indicating the existence of full-length sequences even after 799 synthetic cycles (Fig. S2†).

The last nucleotide was introduced with PTP on the MerMade 6 synthesizer, which also tagged the full-length sequences with a methacrylamide group. Details are given in the ESI.† For deprotection and cleavage, the glass wool was first treated with 10% DBU in ACN, which removed the 2-cyanoethyl groups. Treating with concentrated  $\text{NH}_4\text{OH}$  under typical oligo deprotection conditions gave a mixture of 5'-tagged full-length oligo and un-tagged failure sequences as well as other impurities (Scheme 1). CBP purification was then carried out by co-polymerizing the tagged full-length oligo into a polyacrylamide gel. The failure oligos and many other types of impurities were removed by washing. This gave only the full-length oligo on the polymer. The full-length oligo was then cleaved from the gel using 80% AcOH. After removing the acid, the oligo may be precipitated with  $n\text{BuOH}$  from an  $\text{NH}_4\text{OH}$  solution. This is important for avoiding oligo damage by residue acid if the oligo needs to be stored before use. Otherwise, precipitation may be omitted. The quantity of the oligo was determined to be 27.4  $\mu\text{g}$  (111 pmol) for the synthesis involving 30 mg glass wool. The overall yield for the entire 800-mer synthesis and purification was 0.38% (entry 8, Table 1).

### Characterization of the 800-mer oligo synthesized on glass wool

The 800-mer oligo was characterized with Sanger sequencing. For this purpose, a portion (30 ng) of the CBP-purified 800-mer gene was PCR amplified using high fidelity DNA polymerase. The product was analyzed with agarose gel electrophoresis. As shown in Fig. 1, a band corresponding to 800-mer can be clearly observed. A portion of the PCR product was ligated into the pCR™4Blunt-TOPO™ vector and transformed into chemically competent *E. coli* cells. The transformed cells were grown on agar plates. Colony PCR was performed on selected cell colonies. The PCR products were analyzed with agarose gel electrophoresis. As shown in Fig. 2A–C, all the 48 colonies

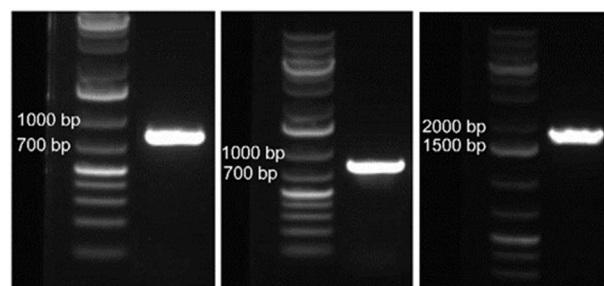


Fig. 1 Gel electrophoresis images of PCR products using CBP-purified oligos, 800-mer synthesized on glass wool (left), 800-mer synthesized on glass beads (middle), and 1728-mer synthesized on glass beads (right), as the template.

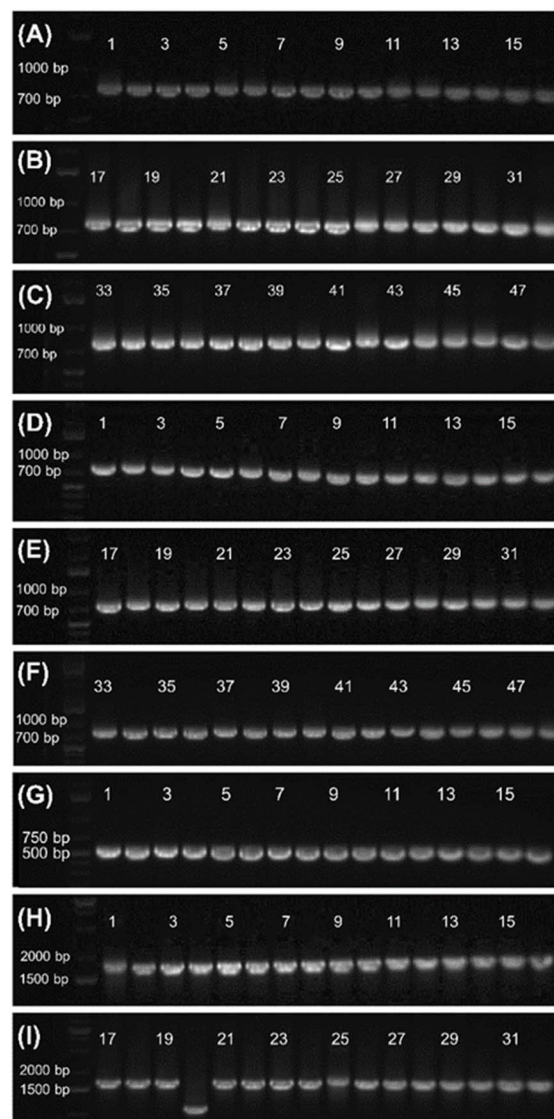


Fig. 2 Gel electrophoresis images of colony PCR products. (A–C) Originated from 800-mer synthesized on glass wool. (D–F) Originated from 800-mer synthesized on glass beads. (G) Originated from 1728-mer synthesized on glass beads. The primers for colony PCR only covered 600 nt of the oligo. (H and I) Originated from 1728-mer synthesized on glass beads. The primers for colony PCR covered the entire oligo. For all the colonies analyzed, all except for lane 20 of (I) contained the anticipated band, evidence of the high reliability of the long oligo synthesis method.



Table 2 Summary of sequencing results

| Entry | Oligo sample                                | 800-mer from glass wool    | 800-mer from glass beads  | 1728-mer from glass beads | 1 <sup>st</sup> 1000 nt of the 1728-mer | Literature error rates <sup>a</sup> |
|-------|---------------------------------------------|----------------------------|---------------------------|---------------------------|-----------------------------------------|-------------------------------------|
| 1     | Total colonies sequenced                    | 48                         | 47                        | 16                        | 16                                      |                                     |
| 2     | Colonies with the correct sequence          | 41                         | 45                        | 7                         | 14                                      |                                     |
| 3     | Rate of the correct sequence                | 85%                        | 96%                       | 44%                       | 88%                                     |                                     |
| 4     | G-to-A substitution/error rate <sup>b</sup> | 0                          | 0                         | 0                         | 0                                       | 0.11% <sup>c</sup>                  |
| 5     | G-to-T substitution/error rate              | 0                          | 0                         | 3/0.0109%                 | 1/0.0063%                               | 0.03%                               |
| 6     | C-to-T substitution/error rate              | 1/0.0026%                  | 0                         | 0                         | 0                                       | 0.02%                               |
| 7     | T-to-C substitution/error rate              | 1/0.0026%                  | 0                         | 1/0.0036%                 | 0                                       | 0.01%                               |
| 8     | A-to-G substitution/error rate              | 0                          | 0                         | 1/0.0036%                 | 0                                       | 0.01%                               |
| 9     | A-to-T substitution/error rate              | 1/0.0026%                  | 0                         | 0                         | 0                                       | <0.01%                              |
| 10    | Single nt deletion/error rate               | 4/0.0104%                  | 1/0.0027%                 | 3/0.0109%                 | 1/0.0063%                               | 0.4% <sup>d</sup>                   |
| 11    | Block deletion/error rate                   | One 10 nt deletion/0.0026% | One 2 nt deletion/0.0027% | Two 2 nt deletion/0.0072% | 0                                       | No data                             |
| 12    | Single nt insertion/error rate              | 0                          | 0                         | 2/0.0072%                 | 0                                       | 0.00–0.01% <sup>e</sup>             |
| 13    | Total error rate <sup>f</sup>               | 0.0208%                    | 0.0054%                   | 0.0434%                   | 0.0126%                                 | 0.58%                               |

<sup>a</sup> Data were from sequencing the 20<sup>th</sup> to 48<sup>th</sup> nucleotide region of chemically synthesized 85-mers. Oligo synthesis conditions: activation, 1*H*-tetrazole in ACN; capping, Ac<sub>2</sub>O in THF, 10% 1-methylimidazole in 10% pyridine/THF; oxidation, 0.02 M I<sub>2</sub> in THF/pyridine/H<sub>2</sub>O; deblocking, 3% TCA in DCM. For more details, see ref. 31. <sup>b</sup> The error rates were calculated by dividing the number of errors by the total number of nucleotides subjected to sequencing. For example, for the 1728-mer synthesized on glass beads, a total of three G-to-A substitutions were found in the data of sequencing 16 colonies; the error rate is 3 ÷ (1728 × 16) = 0.0109%. <sup>c</sup> When DCI was used as the activator, the error rate was lower. <sup>d</sup> 0.1% for each nucleotide. <sup>e</sup> dA 0.005%, dC 0.003%, dG 0.008%, T 0.002%. <sup>f</sup> The total error rate is the sum of individual error rates. It does not represent the probability for a specific nucleotide position in a sequence to have substitution, deletion, addition and other errors.

contained an expected band. Plasmids of the colonies were subjected to Sanger sequencing. The alignment of the sequencing data with the reference sequence is provided in the ESI.† The results are summarized in Table 2. Among the 48 colonies sequenced, 41 contained the correct sequence, corresponding to a success rate of 85% (entries 1–3). The errors in the incorrect sequences include three substitutions, four single nucleotide deletions and one 10 nt deletion (entries 6, 7 and 9–11). The rates for the different errors were all lower than 0.003% except for single nucleotide deletion, which had a rate of 0.0104%. The sum of the error rates was 0.0208% (entry 13).

### Synthesis of an 800-mer oligo on glass beads and oligo characterization

Although the loading of glass beads is predicted to be ~4.8 times lower than that of glass wool, the quantity of oligos produced on them under typical small scale synthesis conditions is still predicted to be higher than 1 pmol, a quantity sufficient for most biological applications.<sup>23,24</sup> For example, with 100 mg glass beads with a loading of 256 nmol g<sup>−1</sup>, assuming an average stepwise yield of 99.7%, which corresponds to an overall yield of 0.25% for a 2000-mer synthesis, the quantity of full-length oligo is ~64 pmol. Considering that glass beads are easier to handle and less likely to generate fine particles that may block the synthesizer, we decided to test long oligo synthesis on glass beads for the purpose of comparison with glass wool. The glass beads were functionalized using the same procedure for functionalizing glass wool (Scheme 2). The loading was determined to be 256 nmol g<sup>−1</sup> (entry 6, Table 1), which is close to the calculated value of 249 nmol g<sup>−1</sup> (entry 4).

Oligo synthesis was conducted under the same conditions using glass wool as the support. The scale was 12.8 nmol, for which 50 mg glass beads were used. Deprotection and cleavage as well as CBP purification were also the same except that only 20 mg (theoretically 5.12 nmol oligo) glass beads were used. The quantity of the oligo obtained was determined to be 168 ng (0.68 pmol) for the synthesis involving 20 mg glass beads. The overall yield for the entire 800-mer synthesis and purification was 0.013% (entry 8, Table 1), which is lower than 0.38% for glass wool. The reason is unclear but may be attributable to the loss of materials in the deprotection and purification process probably due to the increased difficulty to handle smaller quantities of oligos.

The CBP purified 800-mer was also subjected to PCR, cloning and Sanger sequencing. The image of the gel for electrophoresis analysis of the PCR product is shown in Fig. 1. Even though the quantity of oligos was much lower, the band corresponding to the 800-mer is clear. The image of the gel for analysis of colony PCR products is shown in Fig. 2D–F. As can be seen, all colonies selected for the analysis had the 800-mer sequence. Plasmids of 47 colonies were subjected to Sanger sequencing. The data are provided in the ESI.† The results are summarized in Table 2. Among the 47 colonies sequenced, 45 contained the correct sequence, which was 96% (entries 1–3). The errors in the incorrect sequences only include one deletion and one 2 nt deletion (entries 10 and 11). The rates for both errors were 0.0027%. The sum of the error rates was 0.0054% (entry 13).

## Synthesis of a 1728-mer oligo on glass beads and oligo characterization

Encouraged by the success of the 800-mer syntheses, considering that the CBP method can potentially pick up oligos with unlimited length from a crude mixture with unlimited complexity, we decided to synthesize the 1728-mer  $\phi$ 29 DNA polymerase gene. Because glass beads had close to zero errors, and had a higher percentage of correct sequences than glass wool (96% vs. 85%) for the 800-mer syntheses, glass beads were chosen for the synthesis. The synthesis, deprotection, and CBP purification procedures as well as PCR, cloning, and sequencing were the same as described for glass wool. The scale of the synthesis was 33.28 nmol, which corresponds to 130 mg glass beads. The quantity of oligo obtained was 2.83  $\mu$ g (5.28 pmol) for the synthesis using 130 mg glass beads (41 pmol g<sup>-1</sup>, entry 9, Table 1). The overall yield for the entire 1728-mer synthesis and purification was 0.016% (entry 10), which is similar to that of 800-mer synthesis using glass beads as the support.

The image of the gel for electrophoresis analysis of the PCR product of the CBP purified 1728-mer is shown in Fig. 1. As can be seen, the expected band can be clearly observed. Colony PCR was first conducted on 16 colonies using primers targeting only a portion of the 1728-mer (see the ESI†). All colonies were found to have the gene (Fig. 2G). Plasmids of the 16 colonies were subjected to Sanger sequencing. Sequencing data are provided in the ESI,† and the results are summarized in Table 2. Among the 16 colonies sequenced, 7 contained the correct sequence, which corresponds to a success rate of 44% (entries 1–3). The errors in the incorrect sequences include five substitutions, three single nucleotide deletion, two 2 nt deletion and 2 single nucleotide insertion (entries 5, 7, 8, 10 and 11). The sum of the error rates was 0.0434% (entry 13). We also performed gel electrophoresis on colony PCR products of additional colonies using primers covering the entire 1728-mer. Among 32 colonies, only one did not show the expected band (Fig. 2H and I).

## Discussion

For long oligo synthesis, besides the challenge of isolating full-length sequences from a complex mixture, another hurdle is synthesis errors, which include deletions, insertions and substitutions.<sup>31</sup> For the two 800-mer syntheses, the one on glass beads is significantly better (Table 2). For the 47 colonies sequenced, only two errors were found. The total error rate was 0.0054%. The synthesis on glass wool had more errors. The total error rate was 0.0208%, which was about four times higher. The reason may be attributable to the less homogeneous reaction environment during synthesis for glass wool. We did not cut glass wool into short segments because of the concern of generating debris that may block the synthesizer. As a result, the glass wool was not freely movable during reagent delivery. In contrast, the reaction environment for glass beads is more homogeneous. For the synthesis results, it is remarkable that for the first 1000 nt assembly for the 1728-mer synthesis, only two errors were found for the 16 sequences that were sequenced

(Table 2). For nucleotides beyond 1000 nt, the error rate was slightly higher, but it was still surprising low.

Compared with results in the literature using CPG as the support,<sup>31</sup> error rates in the present work were drastically lower (Table 2). For example, among the most frequent substitution errors, which include G-to-A, G-to-T, C-to-T, T-to-C, and A-to-G (entries 4–8),<sup>31</sup> the highest G-to-A substitution was completely eliminated in the present work (entry 4). For all other errors, the rates were also lowered. The sum of the rates of substantial errors for literature syntheses is 0.58%, while that for the present syntheses is less than 0.0434%, which is more than 10 times lower (entry 13). It is noted that the error rates for the present work were from 800-mer and 1728-mer synthesis, while the numbers from the literature was from syntheses of oligos shorter than 100-mer. It is known that error rates increase as oligos grow longer. The increased accuracy of the syntheses on a smooth surface compared with within the pores of CPG may be attributed to higher reaction kinetics in the case of the former. The assumption that reactions on a smooth surface have better kinetics is consistent with discussions in a 1987 patent by Benner.<sup>32</sup>

As mentioned earlier, we successfully synthesized 401-mer and 399-mer oligos on CPG.<sup>10</sup> Compared with that work, the present results are also much better. For the prior study, the bands corresponding to the full-length oligos after PCR amplification of CBP purified oligos were weak (see Fig. 2 in ref. 10) while the bands for the present work are strong (Fig. 1). The gel images of colony PCR results also provided evidence of superiority of the present work. For the prior study, according to gel images, plasmids from 26 out of 64 colonies could be readily estimated not to contain the full-length sequence of the target oligo (see Fig. 3 in ref. 10). For the present work, plasmids from 143 out of 144 colonies that were subjected to the analysis could be estimated to contain the expected sequence. For the prior study, six plasmids that were estimated to contain the full-length sequence based on gel analysis were subjected to Sanger sequencing. Two sequences were correct. Later, we intentionally sequenced 14 additional plasmids that were estimated to contain only a portion of the desired sequence.<sup>33</sup> These sequences were found to contain one or more blocks of deleted nucleotides. The deleted blocks ranged from 8 to over 100 nucleotides. For the 20 sequenced sequences, besides the block deletions, other errors include 11 single nucleotide deletions, and three G-to-A and one T-to-C substitutions. The sequencing data are in the ESI.† Comparing these data with the present ones, it is evident that the major problem for synthesizing long oligos on porous supports is block and single nucleotide deletions. In addition, G-to-A substitution is much more likely to occur with porous supports. The comparison indicates that for long oligo synthesis, supports with a smooth surface should be used.

With the long oligo synthesis results that far exceed the expectations of many researchers including us, one may wonder how this is possible considering the many widely recognized side reactions of oligo synthesis. For example, the acetic anhydride capping efficiency is estimated to be ~90%.<sup>31</sup> Assuming a coupling efficiency of 99%, the deletion sequence in our products would be ~0.1%, far higher than the 0.002–0.01% range we observed (entry 10, Table 2). The total detritylation



time under acidic conditions for the 1728-mer synthesis is  $\sim 47$  hours. Common intuition would suggest high levels of depurination in the products. While answers to these questions are hard to obtain, for the former, it is possible that conducting the synthesis on a smooth surface not only improved the yield of coupling, but also drastically improved the yield of capping. For the latter, occasional exposure of oligos to acid might have less of an effect on depurination than constant acid exposure. It is also possible that depurination may be more likely in the pores than on a smooth surface under the same acidic conditions. Furthermore, it is also possible that depurination might have occurred, but it was not detected by our analysis method. The depurinated oligos were broken under the basic conditions of oligo deprotection and cleavage. The 3'-fragments were washed away during CBP, and the 5'-fragment had to compete with the excess primer for PCR amplification.

It is noted that although the current paper is presented in the context of synthesis of genes, which are double-stranded (ds) DNAs, the long oligo synthesis method can also be used to obtain single-stranded (ss) long oligos or to obtain oligos with site-specific modifications. In these contexts, PCR, cloning and Sanger sequencing should only be carried out using a small portion (*e.g.* about 10 pmol) of the synthesized and CBP purified oligos for the purpose of characterization because PCR and cloning would convert ss oligos to ds oligos, and eliminate the site-specific modifications. The remaining portion (more than 50 pmol) can then be used for the intended applications assuming that the sequence error rates are low, and the errors can be tolerated by the applications. If error-free ss oligos with or without site-specific modifications are required, sequences containing errors could be removed by patching error sites with short oligos and then using immobilized MutS to remove the sequences with errors.<sup>34,35</sup> The short patching oligos in the remaining error-free ss oligos can then be easily removed using size-exclusion filtration, solid phase reversible immobilization (SPRI) bead extraction,<sup>36</sup> or other methods.

## Conclusions

In summary, we demonstrated that direct chemical synthesis of oligos with over 1000 nucleotides can be reliably achieved. The innovations that made this possible include conducting synthesis on a smooth surface, and the use of CBP to isolate full-length sequences from a complex mixture. Conducting synthesis on a surface not only reduced steric hindrance, but also, surprisingly, drastically reduced errors. The longest oligo synthesized is the 1728-mer  $\phi 29$  polymerase gene. Given the high quality of data, we believe that the method can be used to synthesize even longer oligos. Synthesizing 1000-mer oligos has long been a major goal in nucleic acid chemistry. Now that this goal has been achieved, we anticipate that many projects in areas such as synthetic biology, protein engineering, and CRISPR Cas9 gene editing will become easier.

## Data availability

All data are available in the main text or the ESI.†

## Author contributions

Y. Yin (investigation, writing), RA (formal analysis, investigation, writing), Y. Yuan (conceptualization, funding acquisition, formal analysis, project administration, supervision, writing), and SF (conceptualization, funding acquisition, project administration, supervision, writing).

## Conflicts of interest

Michigan Technological University owns the IP associated with the work.

## Acknowledgements

Financial support from NSF (1954041), NIH (GM109288), the Robert and Kathleen Lane Endowed Fellowship (Y. Y.), and the Doctoral Finishing Fellowship (Y. Y.) and assistance from D. W. Seppala (electronics), Z. Song (MS), and A. Galerneau (MS); and NSF equipment grants (2117318 for NMR, 1048655 & 1531454 for MS) are gratefully acknowledged.

## Notes and references

- 1 E. M. Jones, J. P. Marken and P. A. Silver, Synthetic microbiology in sustainability applications, *Nat. Rev. Microbiol.*, 2024, **22**, 345–359, DOI: [10.1038/s41579-023-01007-9](https://doi.org/10.1038/s41579-023-01007-9).
- 2 L. Xu, W. G. Ren, Q. Wang and J. W. Li, Advances in nucleic acid universal influenza vaccines, *Vaccines*, 2024, **12**, 664, DOI: [10.3390/vaccines12060664](https://doi.org/10.3390/vaccines12060664).
- 3 R. Obexer, M. Nassir, E. R. Moody, P. S. Baran and S. L. Lovelock, Modern approaches to therapeutic oligonucleotide manufacturing, *Science*, 2024, **384**, eadl4015, DOI: [10.1126/science.adl4015](https://doi.org/10.1126/science.adl4015).
- 4 L. Villiger, J. Joung, L. Koblan, J. Weissman, O. O. Abudayyeh and J. S. Gootenberg, CRISPR technologies for genome, epigenome and transcriptome editing, *Nat. Rev. Mol. Cell Biol.*, 2024, **25**, 160, DOI: [10.1038/s41580-023-00697-6](https://doi.org/10.1038/s41580-023-00697-6).
- 5 X. Y. Wang, A. N. Li, X. J. Li and H. Y. Cui, Empowering protein engineering through recombination of beneficial substitutions, *Chem.–Eur. J.*, 2024, **30**, e202303889, DOI: [10.1002/chem.202303889](https://doi.org/10.1002/chem.202303889).
- 6 L. Ceze, J. Nivala and K. Strauss, Molecular digital data storage using DNA, *Nat. Rev. Genet.*, 2019, **20**, 456–466, DOI: [10.1038/s41576-019-0125-3](https://doi.org/10.1038/s41576-019-0125-3).
- 7 A. Hoose, R. Vellacott, M. Storch, P. S. Freemont and M. G. Ryadnov, DNA synthesis technologies to close the gene writing gap, *Nat. Rev. Chem.*, 2023, **7**, 144–161, DOI: [10.1038/s41570-022-00456-9](https://doi.org/10.1038/s41570-022-00456-9).
- 8 S. Roux, G. Trubl, D. Goudeau, N. Nath, E. Couradeau, N. A. Ahlgren, Y. C. Zhan, D. Marsan, F. Chen, J. A. Fuhrman, T. R. Northen, M. B. Sullivan, V. I. Rich, R. R. Malmstrom and E. A. Elie-Fadrosh, Optimizing *de novo* genome assembly from PCR-amplified metagenomes, *PeerJ*, 2019, **7**, e6902, DOI: [10.7717/peerj.6902](https://doi.org/10.7717/peerj.6902).



- 9 A. E. Miklos, R. A. Hughes and A. D. Ellington, Design and assembly of large synthetic DNA constructs, *Curr. Protoc. Mol. Biol.*, 2012, DOI: [10.1002/0471142727.mb0323s99](#).
- 10 Y. Yin, R. Arneson, A. Apostle, A. Eriyagama, K. Chillar, E. Burke, M. Jahfetson, Y. Yuan and S. Fang, Long oligodeoxynucleotides: Chemical synthesis, isolation *via* catching-by-polymerization, verification *via* sequencing, and gene expression demonstration, *Beilstein J. Org. Chem.*, 2023, **19**, 1957–1965, DOI: [10.3762/bjoc.19.146](#).
- 11 S. Barthel, S. Palluk, N. J. Hillson, J. D. Keasling and D. H. Arlow, Enhancing terminal deoxynucleotidyl transferase activity on substrates with 3' terminal structures for enzymatic *de novo* DNA synthesis, *Genes*, 2020, **11**, 102, DOI: [10.3390/genes11010102](#).
- 12 M. Eisenstein, Enzymatic DNA synthesis enters new phase, *Nat. Biotechnol.*, 2020, **38**, 1113–1115, DOI: [10.1038/s41587-020-0695-9](#).
- 13 M. A. Jensen and R. W. Davis, Template-independent enzymatic oligonucleotide synthesis (tieos): its history, prospects, and challenges, *Biochemistry*, 2018, **57**, 1821–1832, DOI: [10.1021/acs.biochem.7b00937](#).
- 14 H. H. Lee, R. Kalhor, N. Goela, J. Bolot and G. M. Church, Terminator-free template-independent enzymatic DNA synthesis for digital information storage, *Nat. Commun.*, 2019, **10**, 2383, DOI: [10.1038/s41467-019-10258-1](#).
- 15 M. Pichon and M. Hollenstein, Controlled enzymatic synthesis of oligonucleotides, *Commun. Chem.*, 2024, **7**, 138, DOI: [10.1038/s42004-024-01216-0](#).
- 16 S. Y. Fang and S. Fueangfung, Scalable synthetic oligodeoxynucleotide purification with use of a catching by polymerization, washing, and releasing approach, *Org. Lett.*, 2010, **12**, 3720–3723, DOI: [10.1021/ol101316g](#).
- 17 D. Pokharel and S. Y. Fang, Polymerizable phosphoramidites with an acid-cleavable linker for eco-friendly synthetic oligodeoxynucleotide purification, *Green Chem.*, 2016, **18**, 1125–1136, DOI: [10.1039/c5gc01762a](#).
- 18 D. N. a. M. Eriyagama, S. Shahsavari, B. Halami, B. Y. Lu, F. P. We and S. Y. Fang, Parallel, large-scale, and long synthetic oligodeoxynucleotide purification using the catching full-length sequence by polymerization technique, *Org. Process Res. Dev.*, 2018, **22**, 1282–1288, DOI: [10.1021/acs.oprd.8b00209](#).
- 19 Y. Yin, R. Arneson, A. Apostle, A. M. D. N. Eriyagama, K. Chillar, E. Burke, M. Jahfetson, Y. Yuan and S. Fang, Long oligodeoxynucleotides: chemical synthesis, isolation *via* catching by polymerization, verification *via* sequencing, and gene synthesis and expression demonstration, *ChemRxiv*, 2023, preprint, DOI: [10.26434/chemrxiv-2023-79qj3](#).
- 20 A. P. Guzaev and R. T. Pon, Attachment of nucleosides and other linkers to solid-phase supports for oligonucleotide synthesis, *Curr. Protoc. Nucleic Acid Chem.*, 2013, DOI: [10.1002/0471142700.nc0302s52](#).
- 21 J. M. Palomo, Solid-phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides, *RSC Adv.*, 2014, **4**, 32658–32672, DOI: [10.1039/c4ra02458c](#).
- 22 H. Sugimura, T. Moriguchi, M. Kanda, Y. Sonobayashi, H. M. Nishimura, T. Ichii, K. Murase and S. Kazama, Molecular packing density of a self-assembled monolayer formed from *N*-(2-aminoethyl)-3-aminopropyltriethoxysilane by a vapor phase process, *Chem. Commun.*, 2011, **47**, 8841–8843, DOI: [10.1039/c1cc12541a](#).
- 23 P. Kuhn, K. Wagner, K. Heil, M. Liss and N. Netuschil, Next generation gene synthesis: from microarrays to genomes, *Eng. Life Sci.*, 2017, **17**, 6–13, DOI: [10.1002/elsc.201600121](#).
- 24 H. Kim, J. Jeong and D. Bang, Hierarchical gene synthesis using DNA microchip oligonucleotides, *J. Biotechnol.*, 2011, **151**, 319–324, DOI: [10.1016/j.jbiotec.2011.01.004](#).
- 25 S. Y. Fang and D. E. Bergstrom, Fluoride-cleavable biotinylation phosphoramidite for 5'-end-labeling and affinity purification of synthetic oligonucleotides, *Nucleic Acids Res.*, 2003, **31**, 708–715, DOI: [10.1093/nar/gkg130](#).
- 26 M. H. He, X. S. Wu, S. Mao, P. Haruehanroengra, I. Khan, J. Sheng and M. Royzen, Bio-orthogonal chemistry enables solid phase synthesis and HPLC and gel-free purification of long RNA oligonucleotides, *Chem. Commun.*, 2021, **57**, 4263–4266, DOI: [10.1039/d1cc00096a](#).
- 27 A. Grajkowski, J. Cieslak and S. L. Beaucage, Solid-phase purification of synthetic DNA sequences, *J. Org. Chem.*, 2016, **81**, 6165–6175, DOI: [10.1021/acs.joc.6b01020](#).
- 28 S. Y. Fang and D. E. Bergstrom, Reversible 5'-end biotinylation and affinity purification of synthetic RNA, *Tetrahedron Lett.*, 2004, **45**, 7987–7990, DOI: [10.1016/j.tetlet.2004.09.019](#).
- 29 A. Khanal, F. Long, B. Cao, R. Shahbazian-Yassar and S. Y. Fang, Evidence of splitting 1,2,3-triazole into an alkyne and azide by low mechanical force in the presence of other covalent bonds, *Chem.-Eur. J.*, 2016, **22**, 9760–9767, DOI: [10.1002/chem.201600982](#).
- 30 A. Grajkowski, M. Takahashi, B. M. Cawrse and S. L. Beaucage, An improved PEG-linked solid support for minimizing process-related impurities during solid-phase synthesis of DNA and RNA sequences, *Curr. Protoc.*, 2021, **1**, e108, DOI: [10.1002/cpz1.108](#).
- 31 Y. Masaki, Y. Onishi and K. Seio, Quantification of synthetic errors during chemical synthesis of DNA and its suppression by non-canonical nucleosides, *Sci. Rep.*, 2022, **12**, 12095, DOI: [10.1038/s41598-022-16222-2](#).
- 32 S. A. Benner, Magnetic particles as supports for organic synthesis, *US Pat.*, 4638032, 1987.
- 33 Y. Yin, R. Arneson, Y. Yuan and S. Fang, unpublished data.
- 34 W. Wan, L. Li, Q. Xu, Z. Wang, Y. Yao, R. Wang, J. Zhang, H. Liu, X. Gao and J. Hong, Error removal in microchip-synthesized DNA using immobilized MutS, *Nucleic Acids Res.*, 2014, **42**, e102, DOI: [10.1093/nar/gku405](#).
- 35 N. B. Lubock, D. Zhang, A. M. Sidore, G. M. Church and S. Kosuri, A systematic comparison of error correction enzymes by next-generation sequencing, *Nucleic Acids Res.*, 2017, **45**, 9206–9217, DOI: [10.1093/nar/gkx691](#).
- 36 S. Fang, R. Arneson, Y. Yin and Y. Yuan, *De novo* synthesis of error-free long oligos, *Curr. Protoc.*, 2024, **4**, e70028, DOI: [10.1002/cpz1.70028](#).

