Long Oligos: Direct Chemical Synthesis of Genes with up to 1,728 Nucleotides

Yipeng Yin[a+], Reed Arneson[b+], Yinan Yuan[b*], Shiyue Fang[a*]

[a] Department of Chemistry, and Health Research Institute, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, USA

[b] College of Forest Resources and Environmental Science, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, USA

[*] Emails: yinyuan@mtu.edu and shifang@mtu.edu

[+] Equal contributors

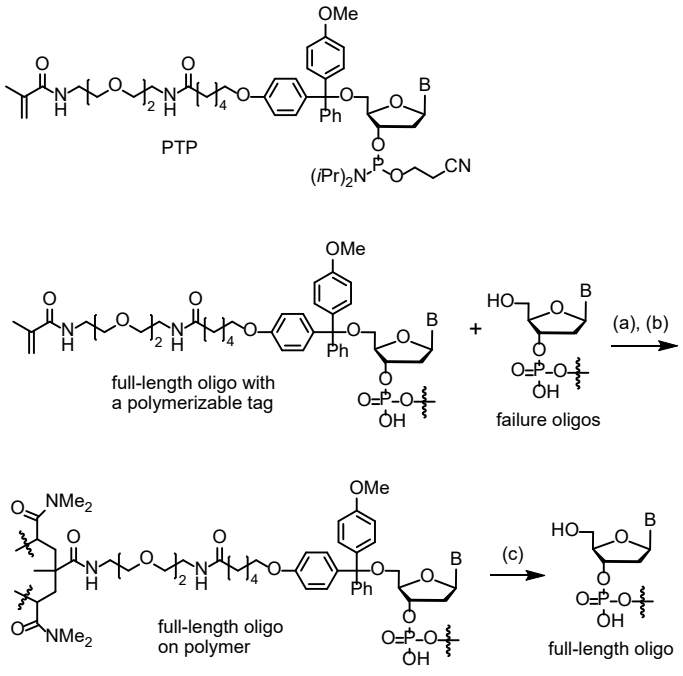
Abstract: The longest oligos that can be chemically synthesized using known methods are typically considered to be 200-mers. Here, we report direct synthesis of an 800-mer green fluorescent protein (GFP) gene and a 1,728-mer Φ29 DNA polymerase gene on an automated synthesizer. Key innovations that enabled the breakthrough include conducting the synthesis on the smooth surface of glass wool or glass bead rather than within the pores of traditional solid supports, and the use of the powerful catching-by-polymerization (CBP) method for the isolation of the full-length oligos from the crude mixture. Conducting the synthesis on smooth surface not only eliminated the steric hindrance that would otherwise prevent long oligo assembly, but also, surprisingly, drastically reduced the errors that commonly occur in traditional oligo synthesis. The long oligos were characterized by cloning followed by Sanger sequencing. We anticipate that the new method for long oligo synthesis will have a significant impact on projects in areas such as synthetic biology, gene editing, protein engineering, and many others.

Introduction

Many areas such as synthetic biology,[1] nucleic acid vaccine,[2] CRISPR Cas9 gene editing,[3] protein engineering,[4] molecular biology, and others require de novo synthesis of single- or double-stranded (ss- or ds-) DNAs longer than 200 nucleotides (nt).[5] Because the state of the art chemical synthesis methods cannot reliably produce oligos longer than 200 nt,[6] these ss- and ds-DNAs have to be produced via biological means such as polymerase chain reaction assembly (PCR assembly) or less commonly ligation using synthetic oligos shorter than 200-mers.[7] While the biological methods have provided the required DNAs for the above mentioned areas to emerge and advance, many sequences desperately needed by some projects are beyond the reach of existing methods.[8] 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, the 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^6A, 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,[9] there is an 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 deoxynucleotidyl transferases (TdT).[6, 10-13] 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 catalytic quantities or recycled. The higher-order structures of long oligos may reduce synthesis efficiency.[13] The coupling time may be lengthy, and the coupling yield may not meet the expectations for typical enzymatic reactions.[13] In addition, enzymatic methods typically lack a capping step, increasing the likelihood of deletion errors.[12] 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.[13] Finally, the method may be difficult to be adapted for synthesizing long oligos with site-specific modifications.



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 × 3. PTP = polymerizable tagging phosphoramidite. B = nucleobase.

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 on the development of a new method called catching-by-polymerization (CBP) for synthetic oligo purification (Scheme 1).[14] 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 with the PTP, and therefore cannot be incorporated into the gel. Oligo purification is then 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 of full-length oligos from the complex mixture generated from thousands of reactions required long oligo synthesis.[15-16] Most recently, using the CBP method, we succeeded in isolation of 400-mers. Sanger sequencing confirmed the sequences.[9] Here, with additional innovations involving the use of glass wool and glass beads as solid support for long oligo synthesis, we report the synthesis of the 800-mer green fluorescent protein (GFP) gene and the 1,728-mer Φ29 DNA polymerase gene, and their isolation with CBP and characterization with Sanger sequencing.

Results and discussion

Previously, we successfully synthesized 400-mer oligos and employed the CBP method to isolate the full-length oligos from the complex mixture of crude products.[9, 17] The key elements that contributed to the success include using controlled pore glass (CPG) with a large pore size of 2,000 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 smooth surface as opposed to within pores of solid supports. A typical concern for this approach is insufficient loading, which may be the major 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 ~8 μm is commercially available and inexpensive. We tested their 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 g/ml, and 3.2 molecules/nm^2 (Table 1),[18] the loading is 1,208 nmol/g (see supporting information 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, a density of 2.2 g/ml, and 3.2 molecules/nm^2, the loading is calculated to be 249 nmol/g. Therefore, the loading of glass wool is ~4.8 times of that of glass beads (Table 1).

Table 1. loading of solid supports, and yields of long oligo synthesis.

Entry	Items	Glass wool[a]	Glass beads	CPG 2K Å tested	CPG 2K Å	Wang resin
1	Density[b]	2.2 g/ml	2.2 g/ml			
2	Diameter[c]	8 µm	58 µm			
3	Loading formula[d]	10.6 ÷ (d × r) µmol/g	15.9 ÷ (d × r) µmol/g			
4	Loading	1.208 µmol/g[e]	0.249 µmol/g[e]	5.405 µmol/g[f]	20-30 µmol/g[g]	0.3-2.5 mmol/g[g]
5	Relative loading	4.8	1	22	~100[h]	~10,000[i]
6	Measured loading[j]	0.981 µmol/g	0.256 µmol/g	5.359 µmol/g		
7	800-mer synthesized[k]	3.7 nmol/g	0.034 nmol/g			
8	800-mer yield	3.7/981 = 0.38%	0.034/256 = 0.013%			
9	1,728-mer synthesized[k]		0.041 nmol/g			
10	1,728-mer yield		0.041/256 = 0.016%			

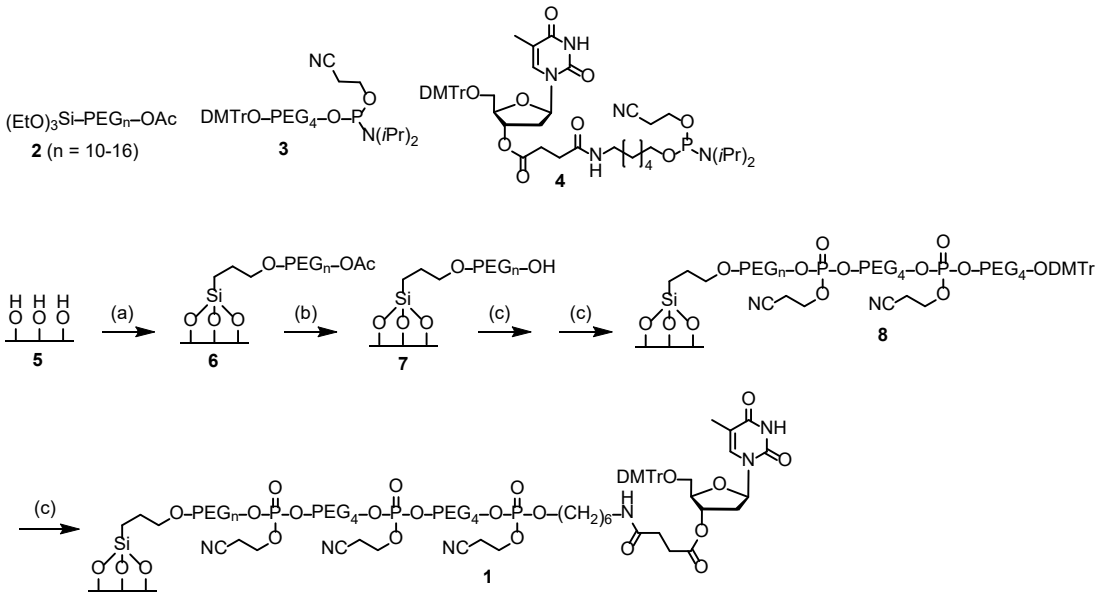
(a) The length of glass wool can be more than 1 cm. Its effect on loading is minimal. (b) Density is that of solid glass. (c)To achieve close to resistance-free flow of liquid, the diameter of glass beads needs to be ~50 µm or larger. (d) The effect of the length of glass wool on loading is minimal and is omitted in the formula. The units for d and r in the formulas are g/mL and µm, respectively. Details for deriving the formulas are in the supporting information. (e) Calculated value assuming 3.2 molecules per nm^2. (f) Given by the manufacturer of the CPG tested. (g) Values from the literature.[19-20] (h) 25 µmol/g is used for the calculation. (i) 2.5 mmol/g is used for the calculation. (j) Measure using trityl assay using literature procedure.[19] (k) Oligo obtained per gram of solid support after CBP purification as determined with Qubit 4 Fluorometer.

For a comparison, the loading of commercial CPG with 2,000 Å diameter is typically 20-30 µmol/50 mg,[19] which is about 100 times higher than that of glass beads (Table 1). The loading of the widely used peptide synthesis Wang resin is 0.3-2.5 mmol/g,[20] 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.[21-22] Using glass wool, with 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 2,000-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 oligos from failure ones. Gel electrophoresis would not be able to resolve either and even if it can be engineered to resolve, the full-length oligo would be invisible on the gel due to its low percentage. Solid phase extraction methods[23-26] may not be suitable for the task as well because the high entropy barrier for reactions between large molecules and solid surface would make the extraction inefficient, and it may be difficult for the large molecules to enter the pores of solid phase in the first place. However, using CBP, the low percentage problem can be overcome. With these considerations, we went ahead to synthesize long oligos on glass wool using the 800-mer GFP gene as the 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-stie (see supporting information and Scheme s1). Glass wool (Figure 1) as depicted

as **5** was activated by treating with piranha solution under previously used conditions.[27] The surface was then PEGylated using **2** to give **6**, the acetyl group of which was removed with ammonia to give **7**.[28] 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 longer reaction time. After two cycles, **8** was given, which was converted to the target functionalized glass wool **1** under similar conditions on the synthesizer using compound **4** as the phosphoramidite monomer in the coupling step. More details for the synthesis are provided in the supporting information. The loading of the glass wool thus functionalized was determined to be 981 nmol/g (Entry 6, Table 1) using a reported method,[19] which is close to the calculated value 1,208 nmol/g (Entry 4).



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_4OH (30%), 55 °C, 2 h. (c) On DNA synthesizer, standard coupling, oxidation and deblocking conditions with modifications; see supporting information for details.

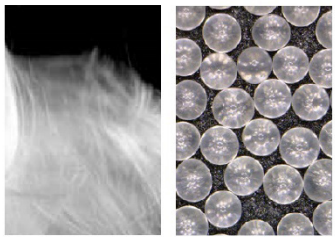


Figure 1. Left, glass wool. Right, glass beads.

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 step of each synthetic cycle in the entire synthesis, the synthesizer manufacturer recommended 1 μmol synthetic cycle was used. While the conditions 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 need reagent to fill pores within the support, than that of traditional supports such as controlled pore glass (CPG), the work was not pursued in the present study. The synthesis was tested on both ABI-394 and MerMade 6 synthesizers. Even though the former consumes more solvents, it was preferred due to its shorter synthesis time. In principle, the synthesis can be carried out consecutively, which would only need for a little more than three days, we paused the synthesis two or more times to refill reagents. With the pauses, the synthesis was completed within four days. Trityl assays 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 MerMade 6 synthesizer. Upon delivery of the deblocking agent, the orange color characteristic of dilute trityl cations could still be observed indicating that the existence of full-length sequences even after 799 synthetic cycles (Figure s1).

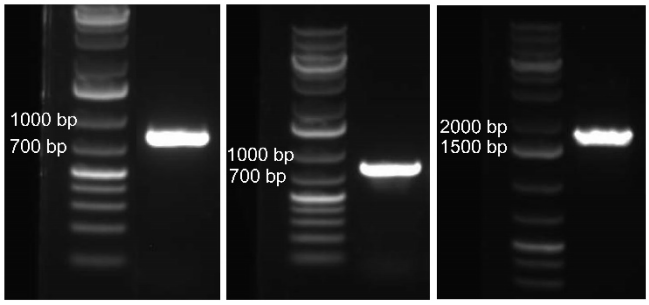


Figure 2. Gel electrophoresis images of PCR products using CBP-purified oligos, 800-mer synthesized on glass wool (left), 800-mer synthesized on glass beads (middle), and 1,728-mer synthesized on glass beads (right), as template.

The last nucleotide was introduced with PTP on the MerMade 6 synthesizer, which also tagged the full-length sequences. Details are given in the supporting information. For deprotection and cleavage, the glass wool was first treated with 10% DBU in ACN, which removed the 2-cyanoethyl groups. Treatment with concentrated NH_4OH 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 (Scheme 1). After removing the acid, the oligo may be precipitated with *n*BuOH from an ammonium hydroxide solution. This is important for avoiding damage of oligo by residue acid if the oligo needs to be stored before its use. Otherwise, precipitation may be omitted. The quantity of the oligo was determined to be 27.4 µg (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).

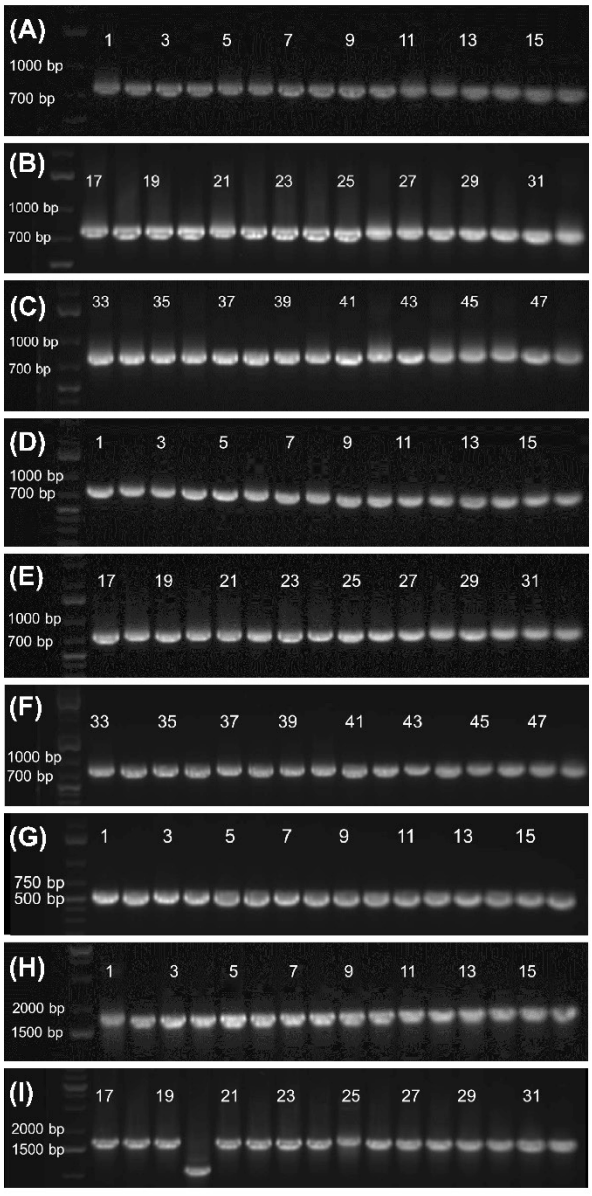


Figure 3. Gel electrophoresis images of colony PCR products. (A-C) Originated from 800-mer synthesized on glass wool. (D-F) Originated from 800-mer synthesized on glass beads. (G) Originated from 1,728-mer synthesized on glass beads. The primers for colony PCR only covered 600-nt of the oligo. (H-I) Originated from 1,728-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, an evidence of the high reliability of the long oligo synthesis method.

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 subjected to PCR amplification using high fidelity DNA polymerase. The PCR product was analyzed with agarose gel electrophoresis. As shown in Figure 2, 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 Figures 3A-C, all the 48 colonies 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 supporting information. The results are summarized in Table 2. Among the 48 colonies sequenced, 41 contained the correct sequence, corresponding to a success rate of 85%. The errors in the incorrect sequences include three substitutions, four singe nucleotide deletions and one 10-nt deletion. 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, Table 2).

Table 2. Summary of sequencing results.

Entry	Oligo sample	800-mer from glass wool	800-mer from glass beads	1,728-mer from glass beads	1st 1,000 nt of the 1,728-mer	Literature error rates[a]
1	Total colonies sequenced	48	47	16	16	
2	Colonies with correct sequence	41	45	7	14	
3	Rate of correct sequence	85%	96%	44%	88%	
4	G-to-A substitution/error rate[b]	0	0	0	0	0.11%[c]
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	Singe nt deletion/error rate	4/0.0104%	1/0.0027%	3/0.0109%	1/0.0063%	0.4%[d]
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%[e]
13	Total error rate[f]	0.0208%	0.0054%	0.0434%	0.0126%	0.58%

(a) Data were from sequencing the 20th to 48th nucleotide region of chemically synthesized 85-mers. Oligo synthesis conditions: activation, 1*H*-tetrazole in ACN; capping, Ac_2O in THF, 10% 1-methylimidazole in 10% pyridine/THF; oxidation, 0.02 M I_2 in THF/pyridine/H_2O; deblocking, 3% TCA in DCM. For more details, see reference.[29] (b) The error rates were calculated by dividing the number of errors by the total number of nucleotides subjected to sequencing. For example, for the 1,728-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 $3 \div (1{,}728 \times 16) = 0.0109\%$. (c) When DCI was used as the activator, the error rate was lower.[29] (d) 0.1% for each nucleotide. (e) dA 0.005%, dC 0.003%, dG 0.008%, T 0.002%. (f) 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.

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 glass wool, the quantity of oligos produced on them under typical small scale synthesis conditions is still predicted to be higher than 1 pmol, the quantity sufficient for most biological applications.[21-22] For example, with 100 mg glass beads with a loading of 256 nmol/g, assuming an average stepwise yield of 99.7%, which corresponds to an overall yield of 0.25% for a 2,000-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 lines of synthesizer, we decided to test long oligo synthesis on glass beads for the purpose of comparison with glass wool. The glass beads (Figure 1) were functionalized using the same procedure for functionalizing glass wool (Scheme 2). The loading was determined to be 256 nmol/g (Entry 6, Table 1), which is close to the calculated value 249 nmol/g (Entry 4).

Oligo synthesis was conducted under the same conditions using glass wool as 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 much lower than the value 0.38% obtained for glass wool. The reason is unclear but may be attributable to the loss of materials in the synthesis, deprotection, cleavage and purification process probably due to the increased difficulty to handle much smaller quantities of oligos.

The CBP purified 800-mer was also subjected to the PCR, cloning and Sanger sequencing process. The image of the gel for electrophoresis analysis of the PCR product is shown in Figure 2. 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 Figures 3D-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 is provided in the supporting information. The results are summarized in Table 2. Among the 47 colonies sequenced, 45 contained the correct sequence, which was 96%. The errors in the incorrect sequences only include one deletion and one 2-nt deletion. The rates for both errors were 0.0027% . The sum of the error rates was 0.0054% (Entry 13).

Synthesis of an 1,728-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 crude mixture with unlimited complexity, we decided to synthesize the 1,728-mer Φ29 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 cleavage, and CBP purification procedures as well as PCR amplification, cloning, and Sanger 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 μg

(5.28 pmol) for the synthesis using 130 mg glass beads (41 pmol/g, Entry 9, Table 1). The overall yield for the entire 1,728-mer synthesis and purification was 0.016% (Entry 10), which is similar to that of 800-mer synthesis using glass beads as support.

The image of the gel for electrophoresis analysis of the PCR product of the CBP purified 1,728-mer is shown in Figure 2. As can be seen, an expected band can be clearly observed. Colony PCR was first conducted on 16 colonies using primers targeting only a portion of the 1,728-mer (see supporting information for details). All colonies were found to have the gene (Figure 3G). Plasmids of the 16 colonies were subjected to Sanger sequencing. Sequencing data are provided in the supporting information, and the results were summarized in Table 2. Among the 16 colonies sequenced, 7 contained the correct sequence, which corresponds to a success rate of 44%. The errors in the incorrect sequences include five substitutions, three single nucleotide deletion, two 2-nt deletion and 2 single nucleotide insertion. The sum of the error rates was 0.0434% (Entry 13). Later, we also performed gel electrophoresis analysis on colony PCR products of additional colonies using primers covering the entire 1,728-mer. Among 32 colonies, only one did not show the expected band (Figures 3H-I).

Additional discussion

For long oligo synthesis, besides the challenge of isolating the low percentage of full-length sequence from the crude mixture, another hurdle is errors in the target sequence, which include deletions, insertions and various substitutions.[29] For the two 800-mer syntheses, the one synthesized on glass beads is significantly better in this aspect (Table 2). For the entire synthesis, only two errors occurred. One was a single nucleotide deletion and the other was a 2-nt deletion. The total error rate was 0.0054%. The synthesis conducted on glass wool had more errors (Table 2), and the total error rate was 0.0208%, which was about four times higher. The reason may be attributable to a less homogeneous reaction environment during automated synthesis for the case of glass wool. We did not cut glass wool to short segments because of the concern of generating small particles that could block the lines of the synthesizer. As a result, the glass wool was not freely movable in the synthesis column during reagent delivery. In contrast, the reaction environment for the case of glass beads is more homogeneous. For the synthesis results, it is remarkable that for the first 1,000-nt assembly for the 1,728-mer synthesis, only two errors were found for the 16 sequences that was sequenced (Table 2). For the nucleotide beyond 1,000-nt, the error rate was slightly higher, but it was still far above our expectations.

Compared with error rates reported in the literature using CPG as the solid support,[29] the 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),[29] the highest G-to-A substitution was completely eliminated in all the three syntheses in the present work (Entry 4). For all other errors, the rates for the present work were consistently lower as well. The sum of the rates of substantial errors for the literature syntheses is 0.58%, while that for the present syntheses is less than 0.0434%, which is more than 10 times lower. It is noted that the numbers for the present work were from 800-mer and 1,728-mer synthesis, while the numbers from the literature was from the synthesis of oligos shorter than 100-mer. It is well known that error rates increase as oligos grow longer.

As mentioned earlier, we successfully synthesized 401-mer and 399-mer oligos on CPG with reduced loading.[9] Compared with results of that work, the present data are also much better. In the prior work, the bands corresponding to the full-length oligos after PCR amplification of the CBP purified oligos were weak (see Figure 2 in the reference[9]) while the bands for the present work are strong (Figure 2). The gel images of colony PCR results also provided evidence of superiority of the present work. For the prior work, according to gel images, plasmids from 23 out of 62 colonies could be readily determined not to contain the sequence of the synthetic oligo or to contain only a portion of it (see Figure 3 in the reference[9]). 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 work, a total of six plasmids that were determined to have the expected sequence based on the gel analysis were subjected to Sanger sequencing. Two sequences were correct, while the other four had at least one error. Later, we intentionally sequenced 14 additional plasmids that were predicted to contain only a portion of desired sequence according to electrophoresis results.[30] We found that these sequences contained 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 deletion errors, other errors include 11 single nucleotide deletions, and three G-to-A and one T-to-C substitutions. The error rates are not calculated because they do not make much sense due to the many deleted blocks of nucleotides. The sequencing data are included in the supporting information. Comparing those data with the present ones, it is evident that the major problem for synthesizing long oligos in pores of CPG and highly cross-linked polystyrene beads is block deletion and single deletion errors. In addition, the G-to-A substitution error is much easier to occur when the synthesis is conducted in pores because in the present work, not even a single G-to-A substitution occurred for all the 111 sequences sequenced. The comparison indicates that for long oligo synthesis, conducting the synthesis on smooth surface should be the choice.

Conclusion

In summary, we have demonstrated that direct chemical synthesis of oligos with over 1,000 nucleotides can be reliably achieved. The innovations that made this possible include conducting synthesis on smooth surface instead of within pores, and isolation of low percentage of full-length oligo from complex crude mixture using the powerful catching-by-polymerization (CBP) method. Conducting oligo synthesis on smooth surface not only reduced steric hindrance, but also drastically reduced errors. The latter is even more important than the former, and was beyond our expectation. The longest oligo synthesized in this work is the 1,728-mer Φ29 DNA polymerase gene. Given the high quality of data as demonstrated by gel electrophoresis and Sanger sequencing, we believe that the method reported here can be used to synthesize even longer oligos. Synthesizing 1,000-mer oligos has long been a major goal in the field of oligo synthesis. Now that this goal has been achieved, we anticipate that many projects in areas such as synthetic biology, protein engineering, CRISPR Cas 9 gene editing, and many others will become easier.

Supporting information

Experimental details, glass wool and glass beads loading calculation, and sequencing results for oligos synthesized on glass wool, glass beads and CPG.

Acknowledgements

Financial support from NSF (1954041), NIH (GM109288), Robert and Kathleen Lane Endowed Fellowship (Y.Y.), and Doctoral Finishing Fellowship (Y.Y.); 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.

References

1. Jones, E. M.; Marken, J. P.; Silver, P. A. Synthetic Microbiology in Sustainability Applications. *Nat. Rev. Microbiol.* **2024**, *22*, 345-359.
2. Xu, L.; Ren, W. G.; Wang, Q.; Li, J. W. Advances in Nucleic Acid Universal Influenza Vaccines. *Vaccines* **2024**, *12*, doi:10.3390/vaccines12060664.
3. Villiger, L.; Joung, J.; Koblan, L.; Weissman, J.; Abudayyeh, O. O.; Gootenberg, J. S. Crispr Technologies for Genome, Epigenome and Transcriptome Editing. *Nat. Rev. Mol. Cell Bio.* **2024**, *25*, 160-160.
4. Wang, X. Y.; Li, A. N.; Li, X. J.; Cui, H. Y. Empowering Protein Engineering through Recombination of Beneficial Substitutions. *Chem. Eur. J.* **2024**, *30*, doi:10.1002/chem.202303889.
5. Ceze, L.; Nivala, J.; Strauss, K. Molecular Digital Data Storage Using DNA. *Nat. Rev. Genet.* **2019**, *20*, 456-466.
6. Hoose, A.; Vellacott, R.; Storch, M.; Freemont, P. S.; Ryadnov, M. G. DNA Synthesis Technologies to Close the Gene Writing Gap. *Nat. Rev. Chem.* **2023**, *7*, 144-161.
7. Roux, S.; Trubl, G.; Goudeau, D.; Nath, N.; Couradeau, E.; Ahlgren, N. A.; Zhan, Y. C.; Marsan, D.; Chen, F.; Fuhrman, J. A.; Northen, T. R.; Sullivan, M. B.; Rich, V. I.; Malmstrom, R. R.; Eloe-Fadrosh, E. A. Optimizing Genome Assembly from Pcr-Amplified Metagenomes. *Peerj* **2019**, *7*, doi:10.7717/peerj.6902.
8. Miklos, A. E.; Hughes, R. A.; Ellington, A. D. Design and Assembly of Large Synthetic DNA Constructs. *Curr. Protoc. Mol. Biol.* **2012**, doi:10.1002/0471142727.mb0471140323s0471142799.
9. Yin, Y.; Arneson, R.; Apostle, A.; Eriyagama, A.; Chillar, K.; Burke, E.; Jahfetson, M.; Yuan, Y.; Fang, S. Long Oligodeoxynucleotides: Chemical Synthesis, Isolation Via Catching-by-Polymerization, Verification Via Sequencing, and Gene Expression Demonstration. *Beilstein J. Org. Chem.* **2023**, *19*, 1957-1965.
10. Lee, H. H.; Kalhor, R.; Goela, N.; Bolot, J.; Church, G. M. Terminator-Free Template-Independent Enzymatic DNA Synthesis for Digital Information Storage. *Nat. Commun.* **2019**, *10*, 2383.
11. Barthel, S.; Palluk, S.; Hillson, N. J.; Keasling, J. D.; Arlow, D. H. Enhancing Terminal Deoxynucleotidyl Transferase Activity on Substrates with 3′ Terminal Structures for Enzymatic De Novo DNA Synthesis. *Genes* **2020**, *11*, 10.3390/genes11010102.
12. Jensen, M. A.; Davis, R. W. Template-Independent Enzymatic Oligonucleotide Synthesis (Tieos): Its History, Prospects, and Challenges. *Biochem.* **2018**, *57*, 1821-1832.
13. Eisenstein, M. Enzymatic DNA Synthesis Enters New Phase. *Nat. Biotechnol.* **2020**, *38*, 1113-1115.
14. Fang, S. Y.; Fueangfung, S. Scalable Synthetic Oligodeoxynucleotide Purification with Use of a Catching by Polymerization, Washing, and Releasing Approach. *Org. Lett.* **2010**, *12*, 3720-3723.
15. Eriyagama, D. N. A. M.; Shahsavari, S.; Halami, B.; Lu, B. Y.; We, F. P.; Fang, S. Y. 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.
16. Pokharel, D.; Fang, S. Y. Polymerizable Phosphoramidites with an Acid-Cleavable Linker for Eco-Friendly Synthetic Oligodeoxynucleotide Purification. *Green Chem.* **2016**, *18*, 1125-1136.
17. Yin, Y.; Arneson, R.; Apostle, A.; Eriyagama, A. M. D. N.; Chillar, K.; Burke, E.; Jahfetson, M.; Yuan, Y.; Fang, S. Long Oligodeoxynucleotides: Chemical Synthesis, Isolation Via Catching by Polymerization,

Verification Via Sequencing, and Gene Synthesis and Expression Demonstration. *ChemRxiv [preprint]* **2023**, 10.26434/chemrxiv-22023-26479qj26433.
18. Sugimura, H.; Moriguchi, T.; Kanda, M.; Sonobayashi, Y.; Nishimura, H. M.; Ichii, T.; Murase, K.; Kazama, S. 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.
19. Guzaev, A. P.; Pon, R. T. Attachment of Nucleosides and Other Linkers to Solid-Phase Supports for Oligonucleotide Synthesis. *Curr. Protoc. Nucleic Acid Chem.* **2013**, *Chapter 3*, 3 2 1-3 2 23.
20. Palomo, J. M. Solid-Phase Peptide Synthesis: An Overview Focused on the Preparation of Biologically Relevant Peptides. *RSC Adv.* **2014**, *4*, 32658-32672.
21. Kuhn, P.; Wagner, K.; Heil, K.; Liss, M.; Netuschil, N. Next Generation Gene Synthesis: From Microarrays to Genomes. *Eng. Life Sci.* **2017**, *17*, 6-13.
22. Kim, H.; Jeong, J.; Bang, D. Hierarchical Gene Synthesis Using DNA Microchip Oligonucleotides. *J. Biotechnol.* **2011**, *151*, 319-324.
23. Fang, S. Y.; Bergstrom, D. E. Fluoride-Cleavable Biotinylation Phosphoramidite for 5′-End-Labeling and Affinity Purification of Synthetic Oligonucleotides. *Nucleic Acids Res.* **2003**, *31*, 708-715.
24. He, M. H.; Wu, X. S.; Mao, S.; Haruehanroengra, P.; Khan, I.; Sheng, J.; Royzen, M. Bio-Orthogonal Chemistry Enables Solid Phase Synthesis and HPLC and Gel-Free Purification of Long RNA Oligonucleotides†. *Chem. Commun.* **2021**, *57*, 4263-4266.
25. Grajkowski, A.; Cieslak, J.; Beaucage, S. L. Solid-Phase Purification of Synthetic DNA Sequences. *J. Org. Chem.* **2016**, *81*, 6165-6175.
26. Fang, S. Y.; Bergstrom, D. E. Reversible 5′-End Biotinylation and Affinity Purification of Synthetic RNA. *Tetrahedron Lett.* **2004**, *45*, 7987-7990.
27. Khanal, A.; Long, F.; Cao, B.; Shahbazian-Yassar, R.; Fang, S. Y. 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.
28. Grajkowski, A.; Takahashi, M.; Cawrse, B. M.; Beaucage, S. L. 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.
29. Masaki, Y.; Onishi, Y.; Seio, K. Quantification of Synthetic Errors during Chemical Synthesis of DNA and Its Suppression by Non-Canonical Nucleosides. *Sci. Rep.* **2022**, *12*, 12095.
30. Yin, Y.; Arneson, R.; Yuan, Y.; Fang, S., unpublished data.