Template for a Protocol Article

ARTICLE TITLE:

De novo synthesis of error-free long oligos

AUTHOR(S) AND CONTACT INFORMATION:

Shiyue Fang*, Department of Chemistry, and Health Research Institute, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, email: shifang@mtu.edu, primary corresponding author

Reed Arneson[+], College of Forest Resources and Environmental Science, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, email: rgarneso@mtu.edu

Yipeng Yin[+], Department of Chemistry, and Health Research Institute, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, email: yipengy@mtu.edu

Yinan Yuan*, College of Forest Resources and Environmental Science, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, email: yinyuan@mtu.edu

* Corresponding authors

[+] Equal contributors

ABSTRACT:

This protocol describes the synthesis of long oligos (up to 401-mer), isolation of the oligos from complex mixture using the catching-by-polymerization (CBP) method, and selection of the error-free sequence via cloning followed by Sanger sequencing. Synthesis of the oligos are achieved under standard automated solid phase synthesis conditions with only minor yet critical adjustments using readily available reagents. The CBP method involves tagging the full-length sequence with a polymerizable tagging phosphoramidite (PTP), co-polymerizing the full-length sequence into a polymer, washing away failure sequences, and cleaving the full-length sequence from the polymer. Cloning and sequencing guided selection of error-free sequence overcome the problems of substitution, deletion, and addition errors that cannot be addressed using any other methods including CBP. Long oligos are needed in many areas such as protein engineering and synthetic biology. The method described here is particularly important for projects that require long oligos containing long repeats or stable higher order structures that are difficult or impossible to produce using any existing technologies.

Basic Protocol 1: Long oligo synthesis
Basic Protocol 2: Catching-by-polymerization (CBP) purification
Basic Protocol 3: Error-free sequence selection via cloning and sequencing
Support Protocol 1: Synthesis of polymerizable tagging phosphoramidite (PTP)
Support Protocol 2: Synthesis of 5'-Bz phosphoramidite

KEYWORDS:

Catching-by-polymerization, cloning, long oligo, sequencing, automated synthesis

INTRODUCTION:

Many research areas such as protein engineering (Kizerwetter et al., 2022; Miklos et al., 2012) and synthetic biology (Garner, 2021; Wurtzel et al., 2019) require long oligos. Currently, oligos up to 200-mer can be synthesized using de novo automated chemical synthesis, but longer ones have to be constructed with molecular biology techniques such as PCR assemble and ligation using short oligos, typically 20- to 60-mers (Eroshenko et al., 2012; Merritt et al., 2014). The PCR and ligation methods can meet most of the needs of long oligos, but they cannot produce difficult sequences such as those containing long repeats, stable secondary structures, and high G-C contents. This protocol describes the synthesis of oligos with length up to 401-mers, the isolation of the full-length oligos from the complex mixture generated by thousands of reactions using the catching-by-polymerization (CBP) technique, and removal of sequences containing substitution, deletion and addition errors via selection of the error-free sequence using the cloning and Sanger sequencing techniques (Yin et al., 2023). A 401-mer segment of the green fluorescent protein (GFP) gene is used as the example (Figure 1), but other sequences can be produced similarly. The protocol provides a means for researchers to access long oligos including those containing long repeats, stable secondary structures, and high G-C contents, and thus are difficult to obtain using existing methods. We anticipate that the long oligo synthesis method will have a significant impact on projects that need long oligos in areas such as protein engineering and synthetic biology.

Basic Protocol 1 describes long oligo synthesis using the 401-mer **1** (see Table 1 for nucleotide sequence) as an example (step 1, Figure 1). Basic Protocol 2 describes the isolation of the full-length sequence using the CBP technique (step 2). During this process, the failure sequences are removed. Basic Protocol 3 describes PCR amplification and cloning of the oligo, which is followed by selecting the correct sequence using gel electrophoresis and Sanger sequencing (steps 3-7). During this process, the sequences containing substitution, deletion and addition errors are removed. The Basic Protocol 1 needs the polymerizable tagging phosphoramidite (PTP) **2** (Figure 1) for attaching the polymerizable methacrylamide group to full-length sequence. Their synthesis is described Support Protocol 1. The Basic Protocol 1 also needs phosphoramidite **3** (Figure 1) for reducing the loading of CPG to reduce the steric hindrance within the pores of CPG during long oligo synthesis. The synthesis of **3** is described in Support Protocol 2.

> *CAUTION:* All experiments must be carried out in a suitable fume hood with efficient ventilation whenever possible. All the materials used in the protocols received little or no biological evaluations and may be harmful/toxic to all or certain individuals. Therefore, lab coat, safety goggles and reagent-impermeable protective gloves must be worn all the time.

BASIC PROTOCOL 1

Basic protocol title:

Long oligo synthesis

Introductory paragraph:

This protocol describes the steps for automated solid phase synthesis, and deprotection and cleavage of long oligos. The synthesis of the *s*ingle-stranded 401-mer fragment (**1**ss, see Table 1 for sequence) of a GFP gene construct on an ABI-394 synthesizer is used as an example. It corresponds to step (1) in Figure 1. The synthesis, deprotection and cleavage are carried out under widely used standard conditions with only slight modifications although the modifications may be crucial for the success of the synthesis of such long oligos. Among the modifications, the last nucleotide at the 5'-end, which is the underlined T in **1**ss (Table 1) in the present example, is incorporated with PTP **2**. This introduces a polymerizable methacrylamide group to the full-length sequence. Because the failure sequences have been capped in each synthetic cycle during the synthesis, PTP **2** does not react with them. After deprotection and cleavage, the crude contains the full-length sequence **4**, which contains a polymerizable methacrylamide group, failure sequences **5**, which do not contain a methacrylamide group, and other impurities (Figure 2). As the synthesis proceeds beyond about 150 nucleotides, the trityl color may be invisible to the naked eye, but the synthesizer's trityl monitor can still reliably monitor the synthesis. All the materials for the protocol except for PTP **2** and phosphoramidite **3** are commercially available. The syntheses of **2** and **3** are described in Support Protocols 1 and 2, respectively.

Materials:

> 5'-DMTr-dA^{Bz}-CE phosphoramidite (cas no. 98796-53-3, Glen Research, cat. no. 10-1000, or Hongene, cat. no. PD1-004)

5’-DMTr-dC^{Ac}-CE phosphoramidite (cas no. 154110-40-4, Glen Research, cat. no. 10-1015, or Hongene, cat. no. PD3-007)
5’-DMTr-dG^{iBu}-CE phosphoramidite (cas no. 93183-15-4, Glen Research, cat. no. 10-1020, or Hongene, cat. no. PD2-004)
5’-DMTr-dT-CE phosphoramidite (cas no. 98796-51-1, Glen Research, cat. no. 10-1030, or Hongene, cat. no. PD4-002)
Polymerizable tagging phosphoramidite (PTP) **2** (synthesized according to Support Protocol 1)
Compound **3** (synthesized according to Support Protocol 2)
Drierite (cas no. 7778-18-9)
Dry acetonitrile (ACN, for dissolving phosphoramidite monomers, Glen Research, SKU 40-4050)
Argon tank (for DNA synthesizer)
Dry ACN (for DNA synthesizer)
Dichloromethane (DCM)
3% Dichloroacetic acid (DCA) in DCM (Glen Research, cat. no. 40-4040)
0.25 M 4,5-Dicyanoimidazole (DCI) in ACN (Glen Research, cat. no. 30-3060)
THF/lutidine/Ac_2O (Cap Mix A) (Glen Research, cat. no. 40-4010)
16% MeIm in THF (Cap Mix B) (Glen Research, cat. no. 40-4220)
0.02 M I_2 in THF/pyridine/H_2O (v/v/v, 70:20:10) (Glen Research, cat. no. 40-4330)
0.2 µmol 2000 Å CPG column (Glen Research, cat. no. 20-2202)
10% DBU in ACN
Nuclease-free water
Saturated NH_4OH (cas no. 1336-21-6)

Bottles that fit DNA synthesizer
Vacuum desiccator
Schlenk gas line with argon source
Rubber septa
Syringes and needles
ABI-394 synthesizer (other synthesizer may be used as well)
1.5 mL microcentrifuge tubes
Shaker
Benchtop centrifuge
Pipette and tips
Vacuum centrifuge evaporator

Protocol steps with *step annotations:*

Synthesize long oligo

1. Dry standard phosphoramidite monomers 5’-DMTr-dA^{Bz}-CE, 5’-DMTr-dC^{Ac}-CE, 5’-DMTr-dG^{iBu}-CE, and 5’-DMTr-dT-CE, and compound **3**, as well as polymerizable tagging phosphoramidite (PTP) **2** in bottles that fit the DNA synthesizer in a desiccator over fresh Drierite under high vacuum overnight.

   Phosphoramidites from commercial sources are usually already dry and packed under argon. Drying them again is optional.

   *Synthesizer manuals typically have information regarding the number of couplings that can be accomplished for a specific synthesis (e.g. 1 µmol synthesis) using a specific quantity of phosphoramidite monomers (e.g. 1 g 5’-DMTr-dT-CE phosphoramidite). The information can be used to estimate the amount of monomers to be dried in this step. For compounds **2** and **3**, they are only needed for one cycle. Quantities in the range of 25 to 50 mg can be sufficient.*

2. Fill the desiccator and the bottles containing the phosphoramidites with argon. Stop the bottles with dry rubber septa.

   The above two steps ensure that the monomers are thoroughly dried, and are under an argon atmosphere.

3. Add dry ACN *via* syringes to the bottles to prepare 0.1 M solutions of the phosphoramidites including the four standard phosphoramidites, and phosphoramidites **2** and **3**.

The ACN used here should be of highest possible quality. It needs to be dry and absent of oxygen as much as possible. The one used in the present example was from Glen Research. See Materials section. It comes in a glass bottle with its mouth stopped with a rubber septa. Do not open the bottle. Instead, connect the bottle via a needle to argon, and use a dry syringe to transfer the ACN into the phosphoramidite bottles.

4. Prepare a solution of 0.05 M **3** and 0.05 M 5'-DMTr-dT-CE phosphoramidite by adding an equal volume of the 0.1 M solutions of the latter to the 0.1 M solution of the former via a dry syringe. Shake gently to mix.
5. Load the following materials needed for oligo synthesis onto the DNA synthesizer: Argon, dry ACN, dry DCM, 3% DCA in DCM, 0.25 M DCI in ACN, THF/lutidine/Ac_2O (Cap Mix A), 16% MeIm in THF (Cap Mix B), and 0.02 M I_2 in THF/pyridine/H_2O (v/v/v, 70:20:10).

   The synthesizer used in the present example was ABI-394. Mermade synthesizer was also tested for the synthesis but it took a much longer time to complete the 401-mer synthesis.

6. Load the solutions of the four standard phosphoramidites, the mixture of **3** and 5'-DMTr-dT-CE phosphoramidite, and compounds **2**, as well as a synthesis column onto the DNA synthesizer.

   Exposure of the solutions to air is unavoidable in this step. However, efforts need to be made to minimize the exposure time as much as possible because the dry ACN has a strong driving force to absorb moisture (as well as oxygen), which can lower the yield of oligo synthesis.

   *In the present example, the 401-mer synthesis takes about 2.5 days. The solution of PTP **2** is used in the last synthesis cycle. It is beneficial to prepare the solution of **2** and load it onto the synthesizer right before it is used. This may minimize decomposition of the compound while staying on the synthesizer for a long period.*

   In the present example, a 0.2 µmol 2000 Å CPG column was used.

7. Perform capping by delivering Cap Mix A and Cap Mix B to the synthesis column for 15 seconds. Wait for 20 minutes.

   This can be accomplished using the synthesizer's cap-to-column function.

   Capping for extended time before long oligo synthesis can potentially prevent oligo synthesis starting from hydroxyl groups resulted from breakage of CPG.

8. Wash away the capping agents by delivering ACN to the synthesis column for ~20 seconds.
9. Perform automated oligo synthesis. Remove the 5'-trityl group at the end of the automated synthesis.

   The present synthesis was carried out using a modified version of the synthesizer manufacturer provided standard 1 µmol synthesis cycle. The reason for using a 1 µmol synthesis cycle for a 0.2 µmol synthesis is to avoid inadvertent insufficient coverage of the CPG by reagents because the 0.2 µmol synthesis cycle uses less reagents. It is possible that using the corresponding modified 0.2 µmol cycle would work equally well. Detritylation: 2% DCA in DCM, 98 seconds. Coupling: 0.1 M phosphoramidite in ACN, 0.25 M DCI in ACN, 2.5 seconds × 2 reagent delivery, 35 seconds waiting. Capping: Cap A, THF/pyridine/Ac_2O; Cap B, 16% 1-methylimidazole in THF, 10 seconds delivery, 10 seconds waiting. Oxidation: 0.02 M I_2 in THF/pyridine/H_2O (70 : 20 : 10, v/v/v), 8 seconds delivery, 15 seconds waiting.

   *To reduce the loading to allow more space for long oligo synthesis, the first dT nucleotide from the 3'-end in the sequence (the bold and italicized T in the present example, see Table 1) should be incorporated using the solution of the mixture of **3** and 5'-DMTr-dT-CE phosphoramidite (instead of the solution of 5'-DMTr-dT-CE phosphoramidite) for the coupling step.*

   Depending on the length of the oligo to be synthesized, one or more reagents may not be enough for the entire synthesis. In that case, the long oligo should be divided into several segments. Synthesize the segment at the 3'-end first. After completion, replenish the reagents, and synthesize the next segment on the solid support carrying the first segment. Repeat until the long oligo synthesis is complete.

   *The last nucleotide at the 5'-end in the long oligo (i.e. the underlined T in the present example, see Table 1) should not be incorporated in this step. Instead, it is introduced during tagging with PTP **2** described in the next steps.*

See Current Protocols articles (Beaucage & Caruthers, 2001; Pon, 2000; Wincott, 2001) for additional details on solid-phase oligonucleotide synthesis.

10. Wash the synthesis column by delivering dry ACN for 15 seconds, and perform a reverse flush for 10 seconds. Repeat the wash for two more times, and then perform a block flush for 5 seconds.

    *All these operations and those in the next steps can be carried out using the functions of the synthesizer manually. Alternatively, they can be carried out automatically by setting up a one cycle synthesis or by embedding the steps at the end of the long oligo synthesis. If the latter is used, do not perform detritylation in the last cycle as that would remove the polymerizable tag introduced by PTP **2**.*

11. Deliver the solutions of PTP **2** and the activator DCI by pressurizing the phosphoramidite and activator bottles using the phosphoramidite-preparation function for 5 seconds, delivering the solutions of **2** and DCI using the base-plus-activator function for 5 seconds, and carrying out the push-to-column function.

    The functions are those of the ABI synthesizer. If other synthesizers are used, use appropriate operations to deliver the reagents so that the solid support is fully covered by the solutions.

12. Wait for 5 minutes.
13. Repeat the above two steps for two additional times.
14. Wash the synthesis column by delivering dry ACN for 15 seconds, and perform a reverse flush for 10 seconds.
15. Perform capping and oxidation under conditions described in the annotation of step 9.

    Capping in this step is optional because there is no additional coupling steps, and thus no possibility of generating deletion sequences. Do not perform detritylation.

16. Wash the synthesis column by delivering dry ACN for 15 seconds, and perform a reverse flush for 10 seconds.
17. Remove the synthesis column from the synthesizer.

    *This gives the long oligo **1**ss that is still fully protected, and linked to CPG.*

Deprotect and cleave long oligo

18. Partition the CPG into two portions and place them in 1.5 mL microcentrifuge tubes. Use one portion for the following steps.

    Partition the CPG into two or more portions is suggested. If an error occurs in the following steps, the remaining portions can be used to correct the errors. Regardless of whether the CPG is partitioned or not, and irrespective of the number of partitions, no adjustments are needed for the following steps.

19. Add 1 mL solution of 10% DBU in ACN.
20. Gently shake the tube at room temperature for 10 minutes.
21. Bring down the CPG by a brief spin.
22. Remove the supernatant with a pipette.
23. Repeat steps 19-22 one time.

    The DBU treatment removes the 2-cyanoethyl protecting groups on the phosphate backbone of the oligo.

24. Wash the CPG with ACN by adding 1 mL of the solvent, shaking gently, spinning briefly, and removing it with a pipette.
25. Wash the CPG with water by adding 1 mL of the solvent, shaking gently, spinning briefly, and removing it with a pipette.
26. Add 0.8 mL saturated NH_4OH. Close the tube tightly. Heat the mixture at 55 ^oC for 16 hours.

    It is important to use fresh saturated NH_4OH. In addition, add while the NH_4OH solution is cold. It may also be helpful to cool the tube containing the CPG on ice before adding.

    The deprotection temperature and time should not deviate from those of suggested conditions significantly. Milder conditions may not remove all protecting groups. Harsher conditions may damage the oligo.

    The NH_4OH treatment removes all remaining protecting groups, and cleaves the oligo from CPG.

27. Cool the mixture to room temperature. Transfer the supernatant to a clean 1.5 mL microcentrifuge tube.
28. Wash the CPG with 0.15 mL water two times. Combine the washes with the supernatant.
29. Evaporate the solution to dryness.

*This gives the crude long oligo **1**ss in its fully deprotected form except that its 5-end is tagged with the polymerizable methacrylamide group. The oligo can be represented with **4** (Figure 2). Besides **4**, the crude contains **5** and other impurities such as small molecules resulted from deprotection.*

BASIC PROTOCOL 2

Basic protocol title:

Catching-by-polymerization (CBP) purification

Introductory paragraph:

This protocol describes the steps for the isolation of the full-length 401-mer **1**ss (see Table 1 for sequence) from the complex mixture resulted from thousands of reactions needed for long oligo synthesis using the CBP method. It corresponds to step (2) in Figure 1. The principle of CBP is depicted in Figure 2. The crude oligo containing **4** and **5** as well as other impurities are mixed with the polymerization monomer *N*,*N*-dimethylacrylamide, and cross-linker *N*,*N*′-methylene-bis(acrylamide). Radical polymerization is then initiated with ammonium persulfate and *N*,*N*,*N*′,*N*′-tetramethylethylenediamine (TMEDA). The full-length sequence **4**, which has a methacrylamide group, is co-polymerized into the polymer giving **6**. The failure sequences **5** as well as other impurities are then washed away. The purified full-length sequence is cleaved from the polyacrylamide gel with an acid by breaking the trityl ether bond in **6**. The CBP method is believed to have unlimited power in terms of the length and quantity of oligos that can be purified. For long oligo synthesis, the quantity of full-length sequence may be very low, but the CBP can pick it out efficiently.

Materials:

Crude long oligo ss**1** (from Basic Protocol 1)

Nuclease-free water
Polymerization solution, containing:
- *N*,*N*-dimethylacrylamide
- *N*,*N*′-methylene-bis(acrylamide)
- Sodium acrylate
- Nuclease-free water
- (See recipe in Reagents and Solutions)

0.23 M ammonium persulfate solution
0.69 M *N*,*N*,*N*′,*N*′-tetramethylethylenediamine (TMEDA) solution
20% NaOAc solution
5% Et_3N solution
Nuclease-free water
80% AcOH
Saturated NH_4OH (cas no. 1336-21-6)
*n*BuOH (cas no. 71-36-3)
10% NH_4OH

1.5 mL microcentrifuge tubes
Benchtop centrifuge
Vortex mixer
50 mL centrifuge tube
Shaker
Pipette and pipette tips

2 mL centrifuge filter unit
Spatula
Vacuum centrifuge evaporator
Qubit fluorometer

Protocol steps with *step annotations:*

Co-polymerize full-length sequence into polymer

1. Dissolve the crude long oligo ss**1**, which can be represented by **4**, in a 1.5 mL microcentrifuge tube in 50 µL water.

   Optionally, the crude can be dissolved in 100 µL, and use only half for the following steps. With this, if an error occurs in the following steps, the remaining half can be used. Regardless of using half or all the crude, no adjustments are needed for the following steps.

2. Add 12 µL polymerization solution.
3. Vortex and spin briefly to mix.
4. Add 5 µL 0.23 M ammonium persulfate solution.
5. Add 5 µL 0.69 M *N*,*N*,*N*',*N*'-tetramethylethylenediamine (TMEDA) solution.
6. Immediately, mix by a brief vortex and spin.

   Ammonium persulfate and TMEDA initiate the acrylamide polymerization reaction.

7. Allow the polymerization to proceed without disturbance at room temperature for 1 hour.

   *Full-length sequence **4** is co-polymerized into a polyacrylamide gel to give **6** (Figure 2). The gel size is less than 100 µL.*

Wash away failure sequences

8. Transfer the gel as one piece into a 50 mL microcentrifuge tube.

   Optionally, the gel can be cut into half or smaller pieces. Smaller pieces can allow more efficient extraction of failure sequences. However, if the pieces are too small, they can be difficult to handle in the next steps.

9. Add 20 mL 20% NaOAc solution.
10. Gently shake the mixture at room temperature overnight.
11. Remove the solution and leave the gel behind.
12. Add 20 mL 5% Et_3N solution.
13. Gently shake the mixture at room temperature overnight.
14. Remove the solution and transfer the gel over the filter of a 2 mL centrifuge filter unit.
15. Grind the gel into small pieces using a spatula.
16. Wash the gel with 0.5 mL water by adding water followed by a spin. Repeat five times.

    The washing procedure might be overly extensive. To save time, a less thorough washing might be sufficient.

    *The use of slightly basic conditions for the overnight washes is to ensure that the trityl linkage in **6** does not break prematurely.*

Cleave full-length sequence from polymer

17. Add a minimum volume of 80% AcOH to the gel over the filter in the centrifuge filter unit.

    The volume should be sufficient to cover the gel. In the present example, 200 µL was used.

18. Incubate with occasional hand shaking at room temperature for 5 minutes.
19. Spin briefly to separate the liquid from the gel. Transfer the filtrate to a 1.5 mL microcentrifuge tube.
20. Dilute the filtrate with 0.8 mL water as soon as possible.

It is important to dilute the filtrate so that the chance of oligo damage by acid is minimized.

21. Repeat steps 17-20 two times.
22. Wash the gel with 0.3 mL water three times. Combine the washes into one 1.5 mL microcentrifuge tube.
23. Stop the bottom of the filtering unit holding the gel with parafilm. Add 0.5 mL water to the gel.
24. Gently shake the mixture at room temperature overnight.
25. Remove the parafilm, spin briefly to separate the liquid from the gel, and transfer the filtrate to a 1.5 mL microcentrifuge tube.
26. Wash the gel with 0.25 mL water two times. Combine the washes with the above filtrate.
27. Evaporate volatiles in the above five microcentrifuge tubes containing the acid extracts and water washes.

    The extracts and water washes may also be combined in one larger tube for evaporation.

28. Combine the residues in the five microcentrifuge tubes into one 1.5 mL microcentrifuge tube by dissolving in water, combining the solutions, and evaporating to dryness.

    *The residue is the CBP purified fully deprotected full-length oligo **1**ss (Table 1). The following steps are optional.*

Precipitate full-length sequence with nBuOH, or co-evaporate with water, to remove residue acid

29. Add 20 µL saturated NH_4OH to the CBP purified oligo **1**ss from step 28.
30. Dissolve the oligo by repeating the process of brief vortex and spin for three or more times.
31. Add 180 µL *n*BuOH. Mix thoroughly by vortex for ~20 seconds.
32. Centrifuge at 14.5k rpm (14.1k × g) for 5 minutes.
33. Remove supernatant carefully so that the precipitated oligo is not disturbed and removed.

    The quantity of the oligo is small, and unlikely to be visible to the naked eye.

34. Evaporate residue liquid left using a vacuum centrifuge evaporator.

    *The residue is the precipitated oligo **1**ss. Its quantity can be measured using a Qubit fluorometer.*

    The above precipitation steps can ensure that any residue acid from the cleavage steps of the CBP procedure is removed.

    The volume of saturated NH_4OH can be more or less than 20 µL. A higher volume can make the dissolution easier but the precipitation may be less efficient resulting in a loss of oligo. The volume of nBuOH is suggested to be nine times of that of saturated NH_4OH.

    If there is a concern of losing oligo, of which quantity is small, the precipitation procedure can be omitted, and the oligo from step 28 can be used directly. However, if the oligo needs to be stored for extended periods, removing residue acids is suggested either using the above precipitation method or using the co-evaporation method described in the next step.

35. Add 30 µL water to the oligo from step 28. Vortex, spin and evaporate the water. Repeat for two more times.

    If there is still a concern of oligo damage by residue acid, replace the above 30 µL water with 10% NH_4OH and co-evaporate for one time. In addition, adding a NaOAc solution with 1 equivalent of sodium cation per phosphate group of the oligo and evaporating to dryness can be more effective. In this case, the cation of phosphate would be exchanged from potentially ammonium cations to sodium, which would not generate any acid. Do not add too much NaOAc because it is a non-volatile salt and is not easy to remove.

    *Whether employing nBuOH precipitation, co-evaporation with water, co-evaporation with 10% NH_4OH, or co-evaporation with a NaOAc solution, the resulting residue is oligo **1**ss, and it is in the form that is more stable and more suitable for storage.*

BASIC PROTOCOL 3

Basic protocol title:

Error-free sequence selection via cloning and sequencing

Introductory paragraph:

This protocol describes the procedure to select error-free sequences from the CBP purified full-length oligos. It corresponds to steps (3-7) of Figure 1. Although the majority of the impurities, which are failure sequences, as well as small molecule impurities are removed during the CBP process, the CBP purified full-length sequences still contain a lot of impurities. They include sequences containing deletion, insertion, and substitution errors. This protocol selects the error-free sequence from the CBP purified single-stranded (ss) full-length sequences by converting them into double-stranded (ds) sequences using PCR, and cloning the ds sequences into a vector. The vector is then transformed into *E. coli* cells. The cells are grown, and multiple cell colonies are chosen for Sanger sequencing of the inserted synthetic sequence. The colonies with the correct inserted sequence are thus identified, from which error-free full-length synthetic sequence can be produced. For the 401-mer synthesis example, three colonies were sequenced, and one of them contain the error-free target sequence.

Materials:

ThermoFisher Phusion™ Hot Start II DNA Polymerase kit (Thermo Fisher Scientific, cat. no. F549S), containing:
- Phusion Hot Start II DNA Polymerase
- 5X Phusion™ High-Fidelity (HF) buffer

NEB deoxyribonucleotide triphosphates (NEB, cat. no. N0447S)
PCR forward primer (p**1a**, custom synthesized by IDT, see Table 1 for sequence)
PCR reverse primer (p**1b**, custom synthesized by IDT, see Table 1 for sequence)
Nuclease free water (IDT, cat. no. 11-04-02-01)
Oligo ss**1** (from Basic Protocols 1 and 2)
TopVision Agarose (Thermo Fisher Scientific, cat. no. R0492)
TAE buffer, containing:
- Tris-base (cas no. 77-86-1)
- Acetic acid (cas no. 64-19-7)
- EDTA disodium salt (cas no. 6381-92-6)

Hydrochloric Acid (cas no. 7647-01-0)
NaOH (cas no. 1310-73-2)
GelRed® 10,000 X in water
DNA gel loading dye (6X) (Thermo Fisher Scientific, cat. no. R0611)
Thermo Fisher Scientific GeneRuler 100 bp Plus ladder, ready-to-use (Thermo Fisher Scientific, cat. no. SM0323)
AMPure XP Bead-Based Reagent (SPRI bead, Beckman Coulter, product no. A63881)
70% ethanol
Zero Blunt™ TOPO™ PCR cloning kit for sequencing (Thermo Fisher Scientific, cat. no. 450031), containing:
- Salt solution (1.2 M NaCl, 0.06 M $MgCl_2$)
- pCR™4-Blunt-TOPO™

NEB® 5-alpha competent *E. coli* (NEB, cat. no. C2987H), containing:
- Competent cells
- SOC media

Kanamycin Sulfate (Thermo Fisher Scientific, cat. no. 11815024)
Luria Broth Base (Invitrogen, cat. no. 12795-027)
Select agar (Invitrogen, cat. no. 30391-023)
Lysis Buffer (see Reagents and Solutions), containing:
- Triton X-100 (cas no. 9002-93-1)
- Tris-base (cas no. 77-86-1)
- EDTA disodium salt (cas no. 6381-92-6)

DreamTaq™ Hot Start Green PCR Master Mix (Thermo Fisher Scientific, cat. no. K9021)
M13 forward primer (p**1c**, custom synthesized by IDT, see Table 1 for sequence)
M13 reverse primer (p**1d**, custom synthesized by IDT, see Table 1 for sequence)

QIAGEN QIAprep Miniprep kit (QIAGEN, cat. no. 27104), containing:
- Buffer P1 with RNase
- Buffer P2
- Buffer N3
- Wash PB
- Wash PE
- Elution buffer (EB)
- miniprep spin column

Qubit™ dsDNA Broad-Range Quantification Assay Kit (Thermo Fisher Scientific, catalog number: Q32853)

Benchmark Scientific MyFuge 12 C1012 Benchtop Centrifuge for Microtubes and PCR tubes (SKU: RS7183)
Ice bucket
Axygen 200 µL PCR tubes (Corning, REF: PCR-0208-FCP-C)
Eppendorf Research Plus P20 pipette (Eppendorf, cat. no. 3123000039) and tips
Eppendorf Research Plus P2.5 pipette (Eppendorf, cat. no. 3123000012) and tips
Applied biosystems 2720 Thermal Cycler (other thermal cyclers should work equally well)
Microwave
125 mL Erlenmeyer flasks
EmbiTec RunOne™ Electrophoresis Cell and gel mold
UV workbench GelDoc-It Imaging System and the VisionWorksLS software
1.5 mL nuclease-free microcentrifuge tubes
Vortex Mixer
Eppendorf Research Plus P200 pipette (Eppendorf, cat. no. 3123000055) and tips
Invitrogen HulaMixer
1.5 mL tube magnetic rack (Sergi Lab Supplies, cat. no. 1005)
-20 freezer
-80 freezer
Water bath that can heat to at least 42 ˚C
Eppendorf Research Plus P1000 pipette (Eppendorf, cat. no. 3123000063)
Lab-Line Orbit Environ-Shaker (for growing bacteria in liquid media)
Sterile petri dishes, 100 mm × 15 mm (for agar plates)
Sterile VWR L-shaped cell spreaders (VWR, cat. no. 89042-018)
Boekel incubator (Boekel, SKU: 132000)
Laminar flow hood
Eppendorf Benchtop Centrifuge 5418
Qubit™ 4 Fluorometer
Computer for data analysis
VWR® 50 mL Standard Line Sterile Centrifuge Tubes with Flat Caps (Catalog number: 10025-698)
pH meter
1 L Media bottles

Protocol steps with *step annotations:*

PCR amplification of long oligo

1. Gently vortex and centrifuge solutions of 5× Phusion™ HF buffer, 10 µM NEB deoxyribonucleotide triphosphates (dNTPs), 10 µM PCR forward primer p**1a**, 10 µM PCR reverse primer p**1b**, and nuclease-free water, and keep them on ice.

   *dNTPs, p**1a**, and p**1b** are in nuclease-free water.*

2. Add 4 µl 5× Phusion™ HF Buffer (1x final concentration) to three separate Axygen 200 µL PCR tubes.

   Multiple PCR reactions for the 401-mer oligo were performed to ensure that there is enough PCR product recovered after SPRI cleaning for the next steps. If more template ODN is used, one PCR reaction might be sufficient too.

3. Add 0.4 µL 10 mM dNTPs solution to the reaction (final concentration of 200 µM for each dNTP).
4. Add 1 µL 10 µM PCR forward primer p**1a** and 1 µL 10 µM PCR reverse primer p**1b** (0.5 µM final concentration for each) to the reaction.
5. Add approximately 20 ng oligo ss**1** to each reaction.
6. Add 0.2 µL 2 U/µL Phusion™ Hot Start II High-Fidelity DNA polymerase to each reaction (0.02 U/µL final concentration).
7. Add nuclease-free water to a final volume of 20 µL to each reaction.

   *The DNA polymerase should be added last. Otherwise, its exonuclease activity may degrade ss**1** and the primers.*

8. Mix the reaction mixture by flicking or pipette mixing to ensure homogeneity. Centrifuge to bring down materials.
9. Place the PCR tube into a thermal cycler.
10. Perform Touch-Down PCR (TD-PCR) amplification under the following conditions: 98 °C for 1 min for initial denaturing; denaturing at 98 °C for 7 sec, annealing at 68.2 °C (decrease annealing temperature by 1 °C for each of the subsequent 9 cycles) for 15 sec, and elongation at 72 °C for 20 sec, for 10 cycles; denaturing at 98 °C for 7 sec, annealing at 58.2 °C for 15 sec, and elongation at 72 °C for 20 sec, for 25 cycles; and a final elongation at 72 °C for 7 min.

11. Add 0.6 g TopVision agarose and 60 ml 1× TAE buffer to a 125 mL Erlenmeyer flask and dissolve by heating in a microwave.

    To avoid overboiling, it is best to microwave in short bursts and swirl the solution after. Repeat until agarose is completely dissolved.

12. Add 6 µL 10,000× GelRed® (1× final solution) to the agarose solution and mix. Pour into an electrophoresis gel mold once it is cooled to the level of not burning upon touching.
13. Place the gel in an electrophoresis tank once fully solidified. Fill the tank with 1× TAE buffer.
14. Take 5 µL PCR product solution from each of the three reactions and mix with 1 µL DNA gel loading dye (6×), separately.
15. Load each of the three mixtures of PCR product and loading dye solutions into a different well of the gel.
16. Load 0.1-0.2 Thermo Fisher Scientific GeneRuler 100 bp Plus ladder, ready-to-use solution into another well of the gel.

    Other ladders that have bands close to the expected size of the PCR product can be used as well.

17. Carry out electrophoresis at 50 volts for 30-45 minutes.
18. Take a picture of the gel using the UV workbench GelDoc-It Imaging System and the VisionWorksLS software.

    Alternative imaging systems and software may be used.

    Gel image is in Figure 3A. Characterization of long oligos are challenging. The image can serve as an indicator if the long oligo synthesis and CBP purification succeed or not. If a band close to the expected location is observed, the results may be positive although at this stage no information about substitution, deletion and addition errors are available. These information will be obtained in the subsequent steps.

Cleaning PCR product using SPRI beads (optional)

19. Combine the remaining 15 µL of solutions from the three Axygen 200 µL PCR tubes into a single 1.5 mL nuclease-free microcentrifuge tubes.

    If gel electrophoresis indicates that one or two PCR reactions are not successful, do not combine them.

20. Resuspend AMPure XP Bead-Based Reagent by gentle vortex or tapping against a counter until the mixture is homogenous throughout.

    AMPure XP Bead-Based Reagent are the SPRI beads. SPRI stands for solid phase reversible immobilization.

21. Add an equal volume of AMPure XP Bead-Based Reagent suspension to the combined PCR production solution at room temperature. Mix by pipetting.

22. Incubate the solution on an Invitrogen HulaMixer at room temperature for 20 minutes.

If an Invitrogen HulaMixer is not available, mixing by manually inverting the tube several times every 3-5 minutes works as well.

23. Remove the tube from the HulaMixer and bring down any solution on the wall of the tube by a brief spin.
24. Place the tube on a 1.5 mL tube magnetic rack and let sit for 3 minutes.
25. Remove the supernatant from the tube and discard it.

It is important not to remove any beads with the supernatant, as that would reduce ODN recovery yield. Should this prove difficult, leave 5 µL supernatant in the tube and proceed to the 70% ethanol wash step.

Do not disturb the beads on the wall of the tube when the tube is on the magnet.

26. Gently rinse the beads with 200 µL 70 % ethanol. Wait for 30 seconds. Remove and discard the ethanol supernatant. Repeat the ethanol wash once.
27. Allow the tube to air dry on the magnetic rack for no more than 2 minutes.

Do not over-dry the beads as that would damage the beads.

28. Remove the tube from the magnetic rack and resuspend the beads in 20 µL nuclease free water.
29. Incubate the solution on an Invitrogen HulaMixer at room temperature for 20 minutes.
30. Remove the tube from the Invitrogen HulaMixer and bring down any solution on the wall of the tube with a brief spin. Place the tube on a 1.5 mL tube magnetic rack and let sit for 3 minutes.
31. Transfer the supernatant containing the purified oligo ds**1** into a clean 1.5 mL nuclease-free microcentrifuge tube.
32. Follow steps 11-18 to perform a gel electrophoresis analysis of the purified oligo ds**1** using 5 µL supernatant.

This step is optional. However, running the gel can determine if SPRI cleaning is carried out successfully before proceeding to the next steps.

Cloning long oligo into vector

33. Add the solution of ds**1** to an Axygen 200 µL PCR tube.

*If the above SPRI cleaning procedure is performed, use 4 µL of the ~15 µL solution. If SPRI cleaning is not performed, follow manufacturer's recommendation and use 0.5-4 µL PCR product solution. Approximately 40 ng of ds**1** is sufficient for successful cloning of the oligo into the vector.*

34. Add 1 µL salt solution from the Zero Blunt™ TOPO™ PCR cloning kit to the tube.

The Zero Blunt™ TOPO™ PCR cloning kit was used because it allows for blunt-end ligation of PCR products, which is what is generated from proofreading enzymes such as the Phusion™ polymerase. If other enzymes are used, such as Taq polymerase, a different kit will be needed since blunt ends are not produced by Taq polymerase.

35. Add 0.5 µL pCR™4-Blunt-TOPO™ solution from the cloning kit to the tube.
36. Add 0.5 µL nuclease free water to the tube
37. Gently mix the solution and incubate at room temperature for 20 minutes.

Use this solution immediately for transformation. Store the remaining solution at -20 ˚C

Transform cells and grow colonies

38. Remove NEB® 5-alpha cells from -80 °C freezer and thaw on ice for 10 minutes.
39. Carefully aliquot 25 µL of cells and place in ice-chilled 1.5 mL nuclease-free centrifuge tube.
40. Add 4 µL solution of the ds**1**- pCR™4-Blunt-TOPO™ ligated vector to the tube, and gently mix by either tapping the tube or swirling with a pipette tip.

Do not pipette mix as it can damage the cells.

41. Incubate the solution on ice for 30 minutes.

42. Incubate the solution in a 42 °C water bath for exactly 30 seconds. Immediately place on ice afterwards and let sit for 5 minutes.
43. Add 250 µL SOC to the tube, and place the tube in a Lab-Line Orbit Environ-Shaker that is moving at 250 rpm and has a temperature of 37 °C.

    To maximize exposure of the surface area of the solution to air, place the tube sideways.

44. Pipette 125 µL cells on an agar plate containing 50 µg/µL kanamycin and spread evenly on the plate using a Sterile VWR L-shaped cell spreaders.

    Kanamycin is the antibiotic suitable for the pCR™4 vector. If a different vector is used, a different antibiotic may be needed.

45. Grow the cells at 37 °C overnight in a Boekel incubator.

Colony PCR

46. Add 50 µL lysis buffer to each of 32 different Axygen 200 µL PCR tubes.
47. Select and number 32 discrete colonies on the agar plate for colony PCR.

    Ensure that each colony chosen appears as a near perfect circle. Otherwise more than one colony may be collected.

48. Using a sterile pipette tip, pick up a portion of a colony and place it in a tube.
49. Use the pipette to eject bacteria from the tip into one of the 32 tubes containing the lysis buffer.

    Only a small portion of the colony should be picked up. The remaining colony will be used for the next steps.

50. Perform the above two steps for each of the 31 remaining colonies.
51. Place the 32 solutions in a thermal cycler and heat at 95 °C for 10 minutes.
52. To a 1.5 mL nuclease-free microcentrifuge tube, add 340 µL DreamTaq™ Hot Start Green PCR Master Mix.

    The DreamTaq™ Hot Start Green PCR Master Mix already has polymerase, buffer, and dNTPs. All that needs to be added is template, primers, and water to the desired final volume.

    This and following three steps are to prepare a Master Mix that can be pipetted into 32 different 200 µL PCR tubes. The quantity prepared here is enough for 34 tubes.

53. To the Master Mix add 34 µL 10 µM solutions of M13 forward (p**1c**) and M13 reverse primers (p**1d**), resulting in a 0.5 µM final concentration for each primer.
54. Add 204 µL nuclease free water to the Master Mix and gently pipette mix.
55. Briefly spin the Master Mix.
56. Add 18 µL Master Mix to each of 32 clean 200 µL PCR tubes.
57. Add 2 µL lysed colonies from step 51 to the correctly numbered tube containing the 18 µL PCR Master Mix from step 56. Do this for all the 32 samples.
58. Gently flick the tubes to mix.
59. Briefly spin the tubes.
60. Place the PCR reaction tubes in a thermal cycler.
61. Perform PCR amplification under the following conditions: 95 °C for 2 minutes for initial denaturing; 33 cycles of denaturing at 95 °C for 30 seconds, annealing at 49 °C for 30 seconds, and elongation at 72 °C for 1 minute; and a final elongation at 72 °C for 7 minutes.
62. Follow steps 11-18 to perform electrophoresis on the PCR products, but only use 5 µL of the PCR reaction and do not mix with DNA gel loading dye (6×).

    No DNA gel loading dye (6×) is added because the DreamTaqTM Hot Start Green PCR Master Mix already contains green loading dye.

    Alternative imaging systems and software may be used.

*Gel image is in Figure 3B. As shown, lanes 2, 6, 7, 9-11, 13, 15, 17-20, 22-24, 27-28, and 30-31 probably have the 401-mer ds**1**. Colonies of these oligos should be selected for sequencing. In this example, colonies 6, 10 and 20 were selected.*

Select correct long oligo via Sanger sequencing

63. Identify the colonies for Sanger sequencing based on the above colony PCR gel analysis.
64. Pick these individual selected colonies into separate standard culture tubes with 5 mL Luria Broth containing 250 µg of kanamycin, and incubate these colonies in a Lab-Line Orbit Environ-Shaker with 225 rpm shaking at 37 °C overnight.

    This can be done using a sterile pipette to pick up the remaining colony from the plate and leaving the tip in the Broth.

65. Next day, stop the shaker and move the colony culture tubes into a laminar flow hood.
66. Transfer about 1.5 mL colony culture from each culture tube into a 1.5 mL nuclease-free microcentrifuge tube. Do this for each selected colony.
67. Centrifuge the tubes at 16,000 × g for 1.5 minutes to precipitate the cells.
68. Discard the supernatants.
69. Repeat steps 66-68 by transferring another 1.5 mL colony culture into the same tube that was used in step 66 for each colony. Repeat until all 5 mL culture of each colony has been spun down.

    The next step is for resuspending the bacterial cells. Freezing and then thawing the cells can make it easier.

70. Resuspend the cell pellet in 250 µL QIAGEN miniprep kit P1 solution. Ensure that no clumps of cells remain. Do this for all selected samples.

    The P1 solution is a resuspension buffer. It contains RNase A, which can degrade all RNAs from bacteria.

71. Add 250 µL QIAGEN P2 solution. Mix by inverting the tube 12 times. Do this for all selected samples.

    The P2 solution is for lysing the cells. As the cells are lysed, the solutions change from cloudy to clear.

72. Incubate all the samples separately at room temperature for 5 minutes.

    It is important not to go over 5 minutes for this step.

73. Add 350 µL QIAGEN N3 and mix by inverting the tube 12 times immediately. Do this for all selected samples.

    Buffer N3 is a neutralization buffer. It needs to be introduced timely so that lysing does not proceed for over 5 minutes.

74. Centrifuge the solutions at 16,000× g for 1 minute.
75. Collect 800 µL supernatant and place in a QIAGEN miniprep spin column. Do this for all selected samples.

    Avoid collecting the precipitate as it would clog the filter in the column.

76. Spin the columns at 16,000 × g for 1 minute. Discard flow-throughs.
77. Add 500 µL QIAGEN PB to each of the columns and spin at 16,000 × g for 1 minute. Discard flow-throughs.

    This wash is optional, but is recommended for bacterial cell-lines that contain endonucleases.

78. Add 750 µL QIAGEN PE to each of the columns and spin at 16,000 × g for 1 minute. Discard flow-throughs.

    Buffer PE is a wash buffer that removes excess salts from the filter.

79. Centrifuge the columns at 16,000× g for 1 minute. Discard flow-throughs.
80. Place the column in a clean 1.5 mL nuclease-free microcentrifuge tube and add 50 µL QIAGEN elution buffer. Do this for all selected samples.

    Add the buffer in the center of the column, where the filter disc is. Do not touch filter disc with the pipette tip as this may tear the filter.

81. Incubate the columns at room temperature for 3 minutes.
82. Centrifuge the column at 16,000 × g for 1 minute to collect the plasmid DNA. Do this for all selected samples.
83. Quantify plasmid DNA with Qubit™ 4 Fluorometer and the Qubit™ dsDNA Broad-Range Quantification Assay Kit. Do this for all the samples.

    The kit has detailed protocols. Other quantification methods can be used as well.

84. Submit the plasmid DNAs for Sanger sequencing using M13 forward (p**1c**) and reverse (p**1d**) primers.

    DNA sequencing services are available in core facilities of Universities or companies.

    *Follow the sample submission guidelines of the sequencing facility to be used for sequencing sample preparation. The present sequencing was conducted at Azenta. For each colony to be sequenced, two samples were submitted. Both contain 400 ng plasmid DNA. One of them was premixed with 25 pmol M13 forward primer (p**1c**). The other was premixed with 25 pmol M13 reverse primer (p**1d**). Both were brought up to a final volume of 15 µL with nuclease-free water.*

85. Assemble the forward and reverse reads into a single contig using appropriate software.

    For the present example, the Cap3 program on command line was used. A web-based version of this software could also be used, and it is available at https://doua.prabi.fr/software/cap3.

86. Align the contigs to the reference sequence using an alignment software.

    The alignment software used for the present example is BLAST at https://blast.ncbi.nlm.nih.gov/Blast.cgi.

    Ensure that all contigs are in the same orientation with the reference so that they can be used in the next step. The orientation can be seen in the "Strand" information of the alignment, and should appear as "Plus/Plus". If it appears as "Plus/Minus", use programs such as that found at https://www.bioinformatics.org/sms/rev_comp.html to generate the reverse complement of the contig.

87. Convene the alignments of all contigs to the reference sequence using the MUSCLE software.

    *The MUSCLE software for this project was used as a command line. It was downloaded through Bioconda. It can also be found at https://www.ebi.ac.uk/jdispatcher/msa/muscle. The alignment result can be found online (Yin et al., 2023). As seen, contigs 6 and 20 had errors. Contig 10 had the correct sequence. Therefore, colony 10 contains the error-free 401-mer oligo **1**. Depending on specific applications, colony 10 can be grown to generate more cells, and oligo **1** or segments of the oligo can be copied using PCR.*

SUPPORT PROTOCOL 1

Support protocol title:

Synthesis of polymerizable tagging phosphoramidite (PTP)

Introductory paragraph:

This Protocol describes the synthesis of the polymerizable tagging phosphoramidite (PTP) **2**. The compound is needed at the end of automated oligo synthesis to attach the polymerizable methacrylamide group to the 5'-end of the full-length sequence for catching-by-polymerization (CBP) oligo purification. The route is shown in Figure 4.

Materials:

4-Hydroxybenzophenone (cas no. 1137-42-4, Aldrich)
Freshly distilled THF (over Na/benzophenone ketal under nitrogen)
THF (not distilled)
4-Methoxyphenylmagnesium bromide solution (0.5 M in THF)
Dry diisopropylamine (distilled over CaH_2 under nitrogen)

Dry ether (distilled over CaH_2 under nitrogen)
Ethyl acetate (EtOAc)
Potassium carbonate (K_2CO_3)
Anhydrous sodium sulfate (Na_2SO_4)
Hexanes
Deuterated chloroform ($CDCl_3$, for NMR)
Potassium phosphate (K_3PO_4)
Methyl-6-bromohexanoate (cas no. 14273-90-6, TCI)
Dimethyl sulfoxide (DMSO)
Dry acetone (dried over anhydrous $MgSO_4$)
Dichloromethane (DCM)
2,2'-(Ethylenedioxy)bis(ethylamine) (cas no. 929-59-9, Aldrich)
Diisopropylethylamide (DIEA)
Methacryloyl chloride
Trifluoroacetic acid anhydride (TFAA)
Drierite
Dry pyridine (distilled over CaH_2 under nitrogen)
Thymidine (cas no. 50-89-5)
Diisopropylammonium tetrazolide (cas no. 93183-36-9, AK Scientific)
Acetonitrile (ACN, distilled over CaH_2 under nitrogen)
N,*N*,*N*′,*N*′-Tetraisopropylphosphorodiamidite (cas no. 102691-36-1)
Triethylamine (Et_3N)
Acetone
Silica gel (for flash chromatography)

Oven (for drying glassware)
2-Neck round-bottom flasks (various sizes)
Magnetic stirring bars
Nitrogen gas line/Schlenk line
Rubber septa
Syringes and needles
Magnetic stirring plates
–20 °C Freezer
Cannula
Needle filter (or cotton and copper wire)
Filter funnel
Filter paper
Rotary evaporator
TLC plate and chamber
Round-bottom flasks
Erlenmeyer flasks
Oil pump
Oil bath
Separatory funnel
Drying tube
Vacuum desiccator
Column (for flash chromatography)
NMR tube
NMR spectrometer
Mass spectrometer

Protocol steps with *step annotations:*

Convert 7 to 8

1. Attached an oven-dried 1 L 2-neck round-bottom flask with a magnetic stirring bar to a nitrogen gas line. The neck that is not attached to the gas line should be stopped with a rubber septa.
2. Add 10 g (50.4 mmol, 1 eq) 4-hydroxybenzophenone (**7**) under positive nitrogen pressure.
3. Add 50 mL dry THF via a syringe. Turn on magnetic stirring.
4. Cool the flask on an ice bath.
5. Add 252.4 mL 4-methoxyphenylmagnesium bromide solution (0.5 M in THF, 126 mmol, 2.5 eq) via a syringe dropwise.
6. Stire the reaction mixture for 3 hours while warming to room temperature gradually.
7. Add 7.28 mL (50.4 mmol, 1 eq) dry diisopropylamine via a syringe. Stir at room temperature for at least minutes.

   *Diisopropylamine quenches excess 4-methoxyphenylmagnesium bromide while compound **8** remains as a dianion. This allows precipitation of **8** with Et_2O.*

8. Add 500 mL dry Et_2O. After mixing, place the reaction flask in an ice bath. Stop magnetic stirring.

   *White precipitate (dianionic **8**) should form.*

9. Store the reaction flask with both necks tightly capped in a –20 °C freezer overnight.
10. Reattach the flask to a nitrogen gas line.
11. Prepare a septa with a cannula inserted through it. Attach a needle filter to the inlet end of the cannula.

    Cotton secured to the inlet end with a copper wire works equally well.

12. Under positive nitrogen pressure, replace the septa on the neck that is not attached to nitrogen with the septa prepared above.
13. Remove the supernatant under positive nitrogen pressure by pushing the inlet end of the cannula below the level of liquid.

    The supernatant contains the anion of diisopropylamine, which is a strong base and may react with water violently. It should be quenched carefully by adding dry ethanol dropwise ideally at 0 °C under nitrogen.

14. Add 10 mL water to the precipitate dropwise under positive nitrogen pressure.
15. Add 200 mL EtOAc.
16. Pour the mixture into a separatory funnel. Partition with 100 mL saturated K_2CO_3.
17. Extract the aqueous layer with 50 mL EtOAc for three times.
18. Combine the organic layer and the extracts. Dry over anhydrous Na_2SO_4. Filter.
19. Concentrate the filtrate to dryness.

    *Compound **8**: 13 g, 84%; red foam; TLC (SiO_2) R_f = 0.2 (3:1 hexanes/EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$) δ 3.82 (s, 3H), 6.77 (d, J = 8.8 Hz, 1H), 6.85 (d, J = 8.7 Hz, 2H), 7.12 (d, J = 8.9 Hz, 2H), 7.19 (d, J = 8.9 Hz, 2H), 7.28-7.32 (m, 7H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 55.3, 81.5, 113.2, 114.7, 114.8, 116.0, 127.1, 127.7, 127.8, 129.1, 129.3, 139.4, 147.2, 154.7, 158.5.*

Convert 8 to 9

20. Attach a 250 mL oven-dried 2-neck round-bottom flask with a magnetic stirring bar to a nitrogen gas line.
21. Add 10 g (32.6 mmol, 1 eq) compound **8**.
22. Add 13.9 g (65.2 mmol, 2 eq) powdered K_3PO_4.
23. Add 6.8 g (32.6 mmol, 1 eq) methyl-6-bromohexanoate.
24. Add 5 mL DMSO via syringe.
25. Add 50 mL dry acetone via a syringe or under positive nitrogen pressure.

    Drying the acetone over anhydrous $MgSO_4$ can be sufficient for the purpose. There is no need for distillation.

26. Stir the reaction mixture vigorously at reflux temperature under nitrogen overnight.
27. Remove acetone under reduced pressure on a rotary evaporator.
28. Remove DMSO under vacuum generated by an oil pump on a rotary evaporator.

29. Partition the residue between the 200 mL EtOAc and 50 mL 5% K_2CO_3.
30. Extract the aqueous layer with 20 mL EtOAc for three times.
31. Combine the organic layer and the extracts. Dry over anhydrous Na_2SO_4. Filter.
32. Evaporate the filtrate to dryness under reduced pressure.
33. Dissolve the residue in minimal DCM. Add hexanes dropwise until the solution becomes cloudy.
34. Add one or more drops of DCM to make the solution clear.
35. Store the mixture at –20 °C overnight.
36. Remove the supernatant giving the product as a yellow oil.
37. Dry the product under high vacuum.

*Compound **9**: 11 g, 77%; light yellow oil; TLC (SiO_2) R_f = 0.4 (3:1 hexanes/EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$) δ 1.49-1.55 (m, 2H), 1.69-1.85 (m, 4H), 2.37 (t, J = 7.5 Hz, 2H), 3.69 (s, 3H), 3.82 (s, 3H), 3.97 (t, J = 6.4 Hz, 2H), 6.84 (t, J = 9.0 Hz, 1H), 7.17-7.20 (m, 4H), 7.28-7.34 (m, 9H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 24.6, 25.6, 28.9, 33.9, 51.5, 55.2, 67.5, 81.3, 113.1, 113.7, 127.0, 127.7, 129.1, 139.3, 139.4, 147.3, 158.1, 158.5, 174.0; HRMS (ESI) m/z calcd for $[M + Na]^+$ $C_{27}H_{30}NaO_5$ 457.1991, found 457.1984.*

Convert 9 to 10

38. Attach a 250 mL 2-neck round-bottom flask with a magnetic stirring bar to a nitrogen gas line.
39. Add 10 g (23.03 mmol, 1 eq) compound **9**.
40. Add 10.2 g (69.09 mmol, 10.1 mL, 3 eq) 2,2'-(ethylenedioxy)bis(ethylamine).
41. Add 0.5 mL water.

Adding too much water will lower reaction efficiency. Not adding water may result in diacylation product (Pappas et al., 2009; Tang & Fang, 2008).

42. Stir the mixture under positive nitrogen flow at room temperature for ~5 minutes.
43. Stir the mixture under nitrogen at 90 °C for 12 hours.
44. Cool the reaction mixture to room temperature under nitrogen.
45. Partition the reaction mixture between 50 mL 10% K_2CO_3 and 50 mL DCM.
46. Extract the aqueous layer with 20 mL DCM for three times.
47. Combine the organic layer and the extracts. Dry over anhydrous Na_2SO_4. Filter.
48. Evaporate the filtrate to dryness under reduced pressure.
49. Dry the product under high vacuum.

*Compound **10**: 10 g, 79%; light yellow thick oil; TLC (SiO_2) R_f = 0.3 (5:2:2:1, Et_2O/ACN/MeOH/Et_3N); ^{1}H NMR (500 MHz, CD_3OD) δ 1.46-1.53 (m, 2H), 1.65-1.80 (m, 3H), 2.23 (t, J = 7.4 Hz, 2H), 2.79 (t, J = 5.2 Hz, 2H), 3.37 (t, J = 5.5 Hz, 2H), 3.50-3.55 (m, 6H), 3.60 (s, 3H), 3.75 (s, 3H), 3.94 (t, J = 6.3 Hz, 2H), 6.81 (t, J = 7.7 Hz, 2H), 7.11-7.14 (m, 4H), 7.21-7.26 (m, 9H); ^{13}C NMR (126 MHz, CD_3OD) δ 25.2, 28.6, 35.5, 38.8, 40.5, 54.3, 67.3, 69.2, 69.9, 71.6, 80.9, 112.5, 113.5, 126.4, 127.2, 127.8, 129.0, 139.8, 139.9, 148.0, 158.0, 158.5, 174.7; HRMS (ESI) m/z calcd for $[M + H]^+$ $C_{32}H_{43}N_2O_6$ 551.3121, found 551.3126.*

Convert 10 to 11

50. Attach a 250 mL oven-dried 2-neck round-bottom flask with a magnetic stirring bar to a nitrogen gas line.
51. Add 10 g (18.1 mmol, 1 eq) compound **10** under positive nitrogen pressure.
52. Add 16.74 mL (90.6 mmol, 5 eq) DIEA via syringe.
53. Add 40 mL dry DCM via a syringe.
54. Turn on magnetic stirring. Cool the mixture to 0 °C.
55. Add 1.94 mL (19.9 mmol, 1.1 eq) methacryloyl chloride via a cannula along the wall of the cooled flask slowly while stirring the reaction mixture vigorously.

Adding slowly along the cooled wall of flask may reduce the likelihood of side reaction involving the trityl hydroxyl group and methacryloyl chloride.

56. Stir the reaction mixture for 12 hours while warming to room temperature gradually.
57. Pour the reaction mixture to a separatory funnel.
58. Partition the reaction mixture with 50 mL 10% K_2CO_3.

59. Extract the aqueous layer with 20 mL DCM for three times.
60. Combine the organic layer and the extracts. Dry over anhydrous Na_2SO_4. Filter.
61. Evaporate the filtrate to dryness under reduced pressure.
62. Dissolve the product in minimal THF containing 1% DIEA.
63. Add hexanes dropwise until the mixture becomes cloudy.
64. Add a few drops of THF to make the solution clear.
65. Store the solution in a –20 °C freezer overnight.
66. Remove the supernatant giving the product as a light yellow oil.
67. Dry the product under high vacuum, giving the product as a white foam.

*Compound **11**: 10 g, 90%, white foam, TLC (SiO_2) R_f = 0.4 (3:2 acetone/5% Et_3N in hexanes); ^1H NMR (500 MHz, $CDCl_3$) δ 1.43-1.50 (m, 2H), 1.64-1.71(m, 2H), 1.73-1.79 (m, 2H), 1.94 (s, 3H), 2.17 (t, J = 7.5 Hz, 2H), 3.39-3.57 (m, 12H), 3.77 (s, 3H), 3.92 (t, J = 6.3 Hz, 2H), 5.29 (d, J = 14.3 Hz, 2H), 5.69 (s, 1H), 6.27 (s, 1H), 6.45 (s, 1H), 6.80 (t, J = 10.3 Hz, 2H), 7.13-7.17 (m, 4H), 7.22-7.28 (m, 9H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 18.7, 25.4, 25.8, 29.0, 36.5, 39.1, 39.4, 55.3, 67.6, 69.8, 70.0, 70.1, 70.2, 81.4, 113.1, 113.6, 119.7, 127.0, 127.8, 129.2, 139.5, 139.9, 147.5, 158.0, 158.5, 168.6, 173.2; HRMS (ESI) m/z calcd for $[M + Na]^+$ $C_{36}H_{46}N_2NaO_7$ 641.3203, found 641.3209.*

Convert 11 to 12

68. Dry 2 g (3.23 mmol, 1 eq) compound **11** with a magnetic stirring bar in a 100 mL 1-neck round-bottom flask in a vacuum desiccator over fresh Drierite under high vacuum overnight.
69. Stop the flask with a rubber septa. Attach the flask to nitrogen gas line via a needle.
70. Add 20 mL dry DCM via a syringe.
71. Cool the flask on an ice bath.
72. Add 1.34 mL (9.70 mmol, 3 eq) trifluoroacetic acid anhydride (TFAA) via a syringe.
73. Stir the reaction mixture for 2 hours while warming to room temperature gradually.

A deep red solution should be formed.

74. Stop magnetic stirring. Remove magnetic stirring bar.
75. Remove DCM on a rotary evaporator under reduced pressure provided by a water aspirator via a drying tube containing Drierite.

It is important to have a drying tube containing Drierite between the water aspirator and the rotary evaporator. Otherwise, the moisture from the aspirator will diffuse to the reaction flask and inactivate the trityl cation intermediate.

76. Remove remaining volatiles including trifluoracetic acid side product and excess TFAA under high vacuum provided by an oil pump.

To protect the oil pump, besides a liquid nitrogen or dry ice clod trap, it is suggested to include a KOH trap between the cold trap and the oil pump.

77. Add a magnetic stirring bar to the flask containing the trityl cation intermediate. Stop the neck with a rubber septa. Attach the flask to a nitrogen gas line via a needle.
78. Add 10 mL dry pyridine via syringe.
79. Dry 0.94 g (3.88 mmol, 1.2 eq) thymidine with a magnetic stirring bar in a 50 mL 1-neck round-bottom flask in a vacuum desiccator over fresh Drierite under high vacuum overnight.

This should be done at the same time of step 68.

80. Stop the flask with a rubber septa. Attach the flask to nitrogen gas line via a needle.
81. Add 10 mL dry DCM via a syringe.
82. Add 2.82 mL (16.17 mmol, 5 eq) DIEA via a syringe.
83. Add 10 mL dry pyridine via a syringe.
84. Stir the reaction mixture at room temperature overnight.
85. Pour the reaction mixture into a separatory funnel containing 50 mL 10% K_2CO_3. Partition the organic and aqueous phases.
86. Extract the aqueous phase with 20 mL DCM for 3 times.

87. Combine the organic layer and the extracts. Dry over anhydrous Na_2SO_4. Filter.
88. Evaporate the filtrate to dryness under reduced pressure.
89. Dissolve the product in minimal DCM containing 1% DIEA.
90. Add hexanes dropwise until the mixture becomes cloudy.
91. Add a few drops of DCM to make the solution clear.
92. Store the solution in a –20 °C freezer overnight.
93. Remove the supernatant giving the product as a light yellow oil.
94. Dry the product under high vacuum, giving the product as a white foam.

*Compound **12**: 2.1 g, 77%, white foam, TLC (SiO_2) R_f = 0.2 (3:2 acetone/hexanes 5% Et_3N); ^1H NMR (500 MHz, $CDCl_3$) δ 1.41 (s, 3H), 1.47-1.77 (m, 6H), 1.95 (s, 3H), 2.20 (t, J = 7.4 Hz, 2H), 2.30-2.41 (m, 2H), 3.37-3.62 (m, 10H), 3.92 (t, J = 6.2 Hz, 2H), 4.08 (s, 1H), 4.57 (s, 1H), 5.32 (s, 2H), 5.71 (s, 2H), 6.42 (t, J = 6.8 Hz, 1H), 6.79-6.84 (m, 4H), 7.14-7.41 (m, 9H), 7.63 (s, 1H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 11.8, 18.6, 25.3, 25.7, 28.9, 36.4, 39.1, 39.4, 40.9, 63.6, 67.6, 69.6, 69.8, 70.1, 70.2, 72.2, 81.3, 84.7, 86.3, 86.8, 111.0, 113.1, 113.2, 113.6, 113.7, 119.7, 127.1, 127.9, 128.1, 129.1, 130.0, 130.1, 135.3, 135.9, 139.4, 139.8, 144.2, 147.4, 150.5, 158.1, 158.6, 164.1, 168.7, 173.2; MS (ESI) m/z calcd $C_{46}H_{58}N_4O_{11}Na$ $[M + Na]^+$ 865.4000, found 865.3998.*

95. Dry 500 mg (0.59 mmol, 1 eq) compound **12** with a magnetic stirring bar in a 100 mL 2-neck round-bottom flask in a vacuum desiccator over fresh Drierite under high vacuum overnight.
96. Fill the desiccator and the flask with nitrogen.
97. Attach one neck of the flask to a nitrogen gas line, and stop the other neck with a rubber septa.
98. Add 150 mg (0.88 mmol, 1.5 eq) diisopropylammonium tetrazolide under positive nitrogen pressure.
99. Add 20 mL dry ACN via a syringe.
100. Add 260 mg (0.88 mmol, 1.5 eq) *N*,*N*,*N*′,*N*′-tetraisopropylphosphorodiamidite via a syringe.
101. Stir the reaction mixture under nitrogen at room temperature overnight.
102. Evaporate the reaction mixture to dryness on a rotary evaporator.
103. Dissolve the residue in the solvent mixture of acetone/hexanes (3:1, v/v) with 5% Et_3N.
104. Load the mixture onto a flash chromatography column (SiO_2), and elute the column with the same solvent system.
105. Combine the fractions containing the pure product under the guidance of TLC R_f = 0.2 (SiO_2, acetone/hexane 3:2, 5% Et_3N).
106. Evaporate volatiles until dryness.

*Compound **2**: 0.51 g, 82%; white foam; TLC R_f = 0.2 (SiO_2, acetone/hexane 3:2, 5% Et_3N); ^1H NMR (500 MHz, $CDCl_3$) δ 1.03-1.05 (m, 3H), 1.15-1.17 (m, 3H), 1.39-1.79 (m, 6H), 1.95 (s, 3H), 2.16 (s, 3H), 2.20 (t, J = 7.5 Hz, 2H), 2.30-2.63 (m, 5H), 3.30-3.62 (m, 10H), 3.93 (s, 2H), 4.18 (s, 1H), 4.67 (s, 1H), 5.31 (s, 1H), 5.70 (s, 1H), 6.32-6.52 (m, 1H), 6.78-6.84 (m, 4H), 7.11-7.40 (m, 9H), 7.64 (s, 1H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 11.1, 11.6, 18.6, 24.5, 25.2, 25.7, 28.9, 36.3, 39.0, 39.4, 45.9, 55.1, 58.1, 63.1, 67.5, 70.0, 73.6, 86.7, 111.2, 113.2, 113.7, 117.4, 119.5, 127.1, 127.9, 130.0, 135.3, 139.9, 144.1, 150.4, 158.0, 158.6, 164.0, 168.5, 173.0; ^{31}P NMR (202 MHz, $CDCl_3$) δ 148.4, 148.7; MS (ESI) m/z calcd for $[M + Na]^+$ $C_{55}H_{75}N_6NaO_{12}P$ 1065.5, found 1065.4.*

SUPPORT PROTOCOL 2

Support protocol title:

Synthesis of 5'-Bz nucleoside phosphoramidite

Introductory paragraph:

This Protocol describes the synthesis of 5'-Bz nucleoside phosphoramidite **3**. The compound is needed for reducing the loading of CPG. The route is shown in Figure 5.

Materials:

Compound **13** (synthesized according to Support Protocol 1)
Thymidine 5'-benzoate (cas no 35898-29-4, Aldrich)
Drierite
Diisopropylammonium tetrazolide (cas no. 93183-36-9, AK Scientific)
Dry acetonitrile (ACN, distilled under CaH_2 under nitrogen)
N,*N*,*N*′,*N*′-Tetraisopropylphosphorodiamidite (cas no. 102691-36-1)

Hexanes
Ethyl acetate (EtOAc)
Silica gel (for flash chromatography)
Deuterated chloroform ($CDCl_3$, for NMR)

2-Neck round-bottom flasks (100 mL)
Magnetic stirring bar
Vacuum desiccator
Oil pump
Nitrogen gas line/Schlenk line
Rubber septa
Syringes and needles
Magnetic stirring plates
1-Neck round-bottom flasks
Rotary evaporator
Erlenmeyer flasks
Column (for flash chromatography)
TLC plate and chamber
NMR tube
NMR spectrometer

Protocol steps with *step annotations:*

1. Dry 2 g (5.77 mmol, 1 eq) compound **13** with a magnetic stirring bar in a 100 mL 2-neck round-bottom flask in a vacuum desiccator over fresh Drierite under high vacuum overnight.
2. Fill the desiccator and the flask with nitrogen.
3. Attach one neck of the flask to a nitrogen gas line, and stop the other neck with a rubber septa.
4. Add 1.48 g (8.66 mmol, 1.5 eq) diisopropylammonium tetrazolide under positive nitrogen pressure.
5. Add 30 mL dry ACN via a syringe.
6. Add 2.61 g (8.66 mmol, 1.5 eq) *N*,*N*,*N*′,*N*′-tetraisopropylphosphorodiamidite via a syringe.
7. Stir the reaction mixture under nitrogen at room temperature overnight.
8. Evaporate the reaction mixture to dryness on a rotary evaporator.
9. Dissolve the residue in the solvent mixture of hexanes/EtOAc (1:1, v/v).
10. Load the mixture onto a flash chromatography column (SiO_2), and elute the column with the same solvent system.
11. Combine the fractions containing the pure product under the guidance of TLC R_f = 0.4 (SiO_2, hexanes/EtOAc 2:3).
12. Evaporate volatiles until dryness.

*Compound **3**: 2.8 g, 89%; white foam; TLC R_f = 0.4 (SiO_2, hexanes/EtOAc 2:3); ^{1}H NMR (500 MHz, $CDCl_3$) δ 1.14-1.26 (m, 12H), 1.61-1.63 (m, 4H), 2.14-2.23 (m, 2H), 2.59-2.64 (m, 2H), 3.58-3.82 (m, 3H), 4.30-4.36 (m, 2H), 4.44-4.66 (m, 1H), 6.31 (s, 1H), 7.20 (d, J = 9.7 Hz, 1H), 7.41 (t, J = 7.5 Hz, 2H), 7.55 (t, J = 7.5 Hz, 1H), 7.80 (d, J = 7.5 Hz, 2H), 8.60 (s, 1H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 12.5, 20.3, 24.5, 39.6, 43.3, 58.1, 63.8, 73.2, 83.5, 84.9, 111.2, 117.7, 123.7, 128.6, 129.4, 133.5, 134.9, 149.5, 150.5, 164.0, 166.0; ^{31}P NMR (202 MHz, $CDCl_3$) δ 148.9, 149.0. The compound is known (Azhayev et al., 1993).*

REAGENTS AND SOLUTIONS:

Polymerization Solution:

- Add 340.5 µL (3.32 mmol) *N*,*N*-dimethylacrylamide.
- Add 17 mg (0.11 mmol) *N*,*N*-methylenebis(acrylamide).
- Add 3 mg (0.032 mmol) sodium acrylate.
- Add 170.5 µL water.
- Mix. Store up to 2 days at 0 °C.

50 µg/µL Kanamycin Solution:

- Add 0.5 g kanamycin to 10 mL sterile water and mix.
- Once dissolved, use a 0.22 µm syringe filter to filter sterilize the solution.
- Aliquot out 1 mL of the solution to 1.5 mL microcentrifuge tubes.
- Store up to 6 months at -20 ˚C.

Agar Plates:

- Add 100 mL nuclease free water to a 1 L media bottle.
- Add 6.25 g Luria Broth base, and dissolve.
- Add 3.75 g of selected agar.
- Microwave the solution in short bursts, swirling between intervals, until the agarose is completely dissolved.
- Add nuclease free water to a final volume of 250 mL.
- Autoclave the media for 30 minutes.
- Allow the media to cool enough so that it can be touched without burning.
- Add 250 µL 50 µg/µL Kanamycin solution and swirl (50 µg/mL of media final concentration).
- Pour approximately 30 mL agar into a petri dish in laminar flow hood.
- Allow the petri dish to sit undisturbed open to air for 30 minutes.
- Cover petri dish with lid and let sit in laminar flow hood overnight.
- Store plates inverted at 8 ˚C. The plates can be used for up to 3 weeks.

Lysis Buffer

- To start, make necessary solutions for preparing the Lysis buffer:
  - 1 M Tris, pH 8.0:
    - Dissolve 60.5 g Tris base in 200 mL distilled water in a 1 L media bottle.
    - Add concentrated hydrochloric acid dropwise until the pH is 8.0.
    - Add distilled water up to 500 mL
    - Autoclave.
    - Keep at room temperature.
  - 0.5 M EDTA
    - Dissolve 18.61 g EDTA disodium salt in 100 mL distilled water.
    - Slowly add NaOH pellets until EDTA dissolves completely (which only happens until pH 8.0).
    - Autoclave.
    - Keep at room temperature.
- For a total of 50 mL lysis buffer, add 20 mL nuclease free water to a 50 mL sterile centrifuge tube.
- Add 0.5 mL Triton X-100 (1% V/V).
- Add 1 mL 1 M Tris, pH 8.0 (20 mM final concentration).
- Add 0.2 mL 0.5 M EDTA (2 mM final concentration).
- Add distilled water up to 50 mL and ensure everything is dissolved.
- Keep at room temperature.

TAE Buffer, 50X:

- Add 500 mL nuclease free water to a beaker.
- Add 242 g Tris-base and dissolve (2 M final concentration).
- Slowly add 57.1 mL 100% acetic acid (1 M final concentration).
- Add 100 mL 0.5 M EDTA (50 mM final concentration).
- Adjust the volume to 1 L with nuclease free water.

- Store solution at room temperature. The buffer can be used for up to 6 months.
- To prepare 500 mL 1× TAE, mix 10 mL 50× TAE stock with 490 mL nuclease free water.

COMMENTARY:

Background Information:

Oligos longer than 200-mer are needed for projects in many areas such as synthetic biology (Garner, 2021; Wurtzel et al., 2019), protein engineering (Kizerwetter et al., 2022; Miklos et al., 2012), CRISPR/Cas9 gene editing (Tyagi et al., 2020), mRNA medicine (Chavda et al., 2022), basic molecular biology and others (Ceze et al., 2019). They are mostly produced via PCR assembly of short oligos with lengths ranging from 20 to 60 nucleotides (Hoose et al., 2023; Ma et al., 2012; Miklos et al., 2012; Roux et al., 2019). Besides issues such as complexity of procedure, and error-proneness, it cannot produce long oligos with difficult elements such as long repeats, stable higher-order structures, and high or low G/C contents. The protocols in this paper is targeted to solve the problem. Instead of assembling short synthetic oligos together using PCR or ligation, the method described here involves synthesizing long oligos such as the example 401-mer (**1**) directly on an automated synthesizer. The synthesis is conducted under primarily standard conditions, but with modifications that may be crucial for success. To reduce steric hindrance, CPG with the largest pore (2,000 Å) available from commercial sources is used, and the loading is reduced to approximately half by using a solution of the mixture of **3** and the standard 5'-DMTr-T-CE phosphoramidite for incorporating a T early in the synthesis. Because the 5'-benzoyl group of **3** is stable in subsequent detritylation conditions, assuming **3** and the standard phosphoramidite have similar reactivity for oligo synthesis, half of the reactive sites on the CPG are blocked. In addition, before the commencement of the synthesis, the CPG is capped for extended periods of time to block any hydroxyl groups resulted from breakage of CPG by mechanical forces from participating in the oligo synthesis. The coupling time is slightly longer than standard conditions to increase synthesis yield. For deprotection, the 2-cyanoethyl groups are removed with DBU before the standard ammonium hydroxide treatment. This is intended to prevent the reaction between the deprotection side product acrylonitrile and oligo.

With all the above means, for a synthesis of oligos longer that 200-mer, the percentage yield of the desired full-length product is still low. For example, for a 400-mer 1 µmol synthesis, assuming an average stepwise yield of 99.0%, the overall yield would be close to 2%, which corresponds to 20 nmol. Although the quantity of the oligo is still sufficient for many biological applications, isolation of the full-length sequence is challenging. For example, HPLC cannot resolve such long oligos. Gel electrophoresis may be able to resolve, but identification of such low percentage of full-length sequence on the gel would be difficult. Increasing gel loading cannot solve the problem because that would lower resolution. Solid phase extraction method can be considered (Fang & Bergstrom, 2003, 2004; Grajkowski et al., 2016; He et al., 2021), but for a long oligo molecule to attach to a solid support, a high entropy barrier must be overcome, and the process is expected to be inefficient. Moreover, the long oligo molecules can be difficult to enter the pores of the solid support.

The protocols described here solve these problems using the catching-by-polymerization (CBP) method (Figure 2) (Eriyagama et al., 2018; Fang & Fueangfung, 2010; Pokharel & Fang, 2016). During oligo synthesis, the last nucleotide at the 5'-end is incorporated using a polymerizable tagging phosphoramidite (PTP, **2**), which carries a polymerizable methacrylamide group, instead of a standard phosphoramidite. Because the failure sequences are capped in each synthetic cycle during the synthesis, they do not react with **2**. After deprotection and cleavage, the crude product contains the tagged **4** and untagged impurities such as **5** and others. The mixture is subjected to the radical acrylamide polymerization conditions. The full-length **4** is co-polymerized into a polyacrylamide gel to give **6**, while **5** and other impurities are in solution, which are washed away. The full-length sequence **1**ss is then cleaved from the gel. Using this method, even if the percentage of full-length sequence is extremely low, it can be isolated. Therefore, the CBP method is well suited for purification of long oligos, where the percentage of desired full-length sequence is almost always extremely low.

The full-length sequence purified by the CBP method is still not pure. It contains impurities such as sequences containing deletion, addition and substitution errors. Possible origins of these errors and their rates of occurrence have been discussed in the literature (Masaki et al., 2022). These impurities in the context of long oligos cannot be removed using methods such as HPLC and gel electrophoresis, as well as solid phase extraction methods (Fang & Bergstrom, 2003, 2004; Grajkowski et al., 2016; He et al., 2021). The present method uses cloning and Sanger sequencing to solve the problem (steps 3-7, Figure 1). Therefore, the CBP purified full-length sequence **1**ss is converted to **1**ds and amplified using PCR.

The PCR product is then cloned into a vector, which is transformed to *E. coli* cells. The cells are grown, and selected colonies are subjected to colony PCR followed by Sanger sequencing. Error-free long oligo can then be produced from the colonies that contain sequencing confirmed sequence.

The error-free long oligo synthesis method is ready to make an impact on many research areas by providing long oligos that contain difficult elements such as long repeats, stable higher-order structures, and high or low GC contents, and are thus inaccessible using any known method. Such oligos are frequently declined by custom synthesis companies even if they are shorter than 200-mer. The areas that the method will make an impact on include synthetic biology, protein engineering, CRISPR/Cas9 gene editing, mRNA medicine, and potentially many others.

Critical Parameters:

For long oligo synthesis, the average stepwise yields (ASWY) must be more than 99%. Usually, at the beginning of each synthesis, the ASWY are lower, and sometimes can be as low as 85%. However, that does not matter. The ASWY goes up gradually as the synthesis proceeds, and usually can reach more than 99.5% when the synthesis reaches about 60-mer. The key for this desired yield to be achieved is to keep the solutions of phosphoramidite monomers and the activator dry and devoid of oxygen as much as possible. The details for achieving this are provided in the protocol and they need to be followed carefully. The quality of phosphoramidite monomers is certainly important. We primarily purchased phosphoramidites from Glen Research or Hongene Biotech. Those from other companies may work equally well. Other factors that are important for the success of long oligo synthesis include the use of CPG with the largest possible pores (2,000 Å), reducing the loading to half, capping the CPG before commencement of the synthesis, and longer than standard coupling times. Details of these are in the protocol.

For oligo deprotection, it is important to remove the 2-cyanoethyl phosphate protecting groups with DBU first because it is predictable that deprotected long oligos are more likely to react with the acrylonitrile deprotection side product than short oligos. For CBP purification, during washing failure sequences, cutting the gel into smaller pieces can increase the efficiency of removing failure sequences, but can make gel transfer in the next steps inconvenient. Appropriate balance should be kept in mind. After cleaving the full-length sequence from the polyacrylamide gel with acid, it is important to dilute the solution as soon as possible to avoid potential damage of oligo under acidic conditions.

Troubleshooting:

Table 2. Troubleshooting Guide

Problem	Possible Cause	Solution
Poor ASWY	Phosphoramidites are not dry. ACN are not dry. Activator is not dry.	Use fresh dry phosphoramidites, activator and solvents. When loading the reagents onto the synthesizer, minimize exposure to air.
No oligo after CBP	Oligo is damaged by acid during cleavage. Oligo is not extracted from polyacrylamide gel. Oligo is lost during *n*BuOH precipitation.	Make sure to dilute the solution of oligo cleaved from the polyacrylamide gel timely. When extracting full-length oligo from polyacrylamide gel, the gel should be grinded to small pieces to increase extraction efficiency. When performing precipitation of oligo using *n*BuOH, the ratio of NH_4OH solution and *n*BuOH should be appropriate. We suggest 9:1. In addition, minimize the total volume of NH_4OH solution and *n*BuOH.
Quantity of oligo is low after PCR amplification	A too small portion of CBP-purified oligo is used as template.	Use more CBP-purified oligo as template for PCR.

Understanding Results:

An average stepwise yield (ASWY) of 99.4% as indicated by the ABI-394 synthesizer after a consecutive 60 synthesis cycles indicates successful synthesis. Before about 60 cycles, the ASWY may be lower, but that should not be a significant concern. Figure 3A is used to evaluate the results of long oligo synthesis and CBP purification. A strong band (lane 1) at the expected location in reference to a ladder suggests success. The band in the present example is relatively weak, but sufficient to indicate positive results. If a stronger band is preferred, a higher quantity of template can be used for the PCR. Figure 3B is for the determination of colonies that contain the full-length oligo. The bands with smallest migration distance such as those in lanes 2, 6 and 7 indicate positive results. Oligos from colonies corresponding to these bands should be selected for Sanger sequencing. The bands with longer migration distances such as those in lanes 1, 3, 4, and 8 indicate negative results, and oligos from colonies corresponding to them may not contain the synthetic oligo or the synthetic oligo may contain multiple deletions.

For the synthesis of phosphoramidites, all the intermediate compounds should appear mainly as one spot on TLC. However, the final phosphoramidites **2** and **3** may appear as two spots under certain TLC conditions due to diastereoisomers originated from the chiral phosphorus atom. This is acceptable and there is no need for separating the two spots. The intermediate compounds should give acceptable ^{1}H and ^{13}C NMR, and expected MS. For phosphoramidites **2** and **3**, ^{1}H and ^{13}C NMR are quite complex. ^{31}P NMR and LRMS are better methods for analysis of the results. One or two peaks in the region around 150 ppm in ^{31}P NMR indicate positive results. Occasionally, small peaks in the region around 0 ppm may appear. These are from the oxidized phosphoramidites. In this case, as long as the peaks are significantly less intense than the peaks in the region around 150 ppm, the materials can be used for oligo synthesis without too much concern.

Time Considerations:

For a synthesis of 400-mer oligo on an ABI-394 synthesizer, it takes two to three days depending on if the synthesis has to be stopped in the middle for refilling reagents. Oligo deprotection and cleavage take one day. We did not test faster deprotection conditions such as 10% DBU in ACN followed by saturated NH_4OH/65 ^{o}C/8 hours, AMA (1:1 v/v saturated NH_4OH/saturated methylamine)/room temperature/2 hours, AMA/37 ^{o}C/30 minutes, or AMA/55 ^{o}C/10 minutes. These faster deprotection conditions may work equally well. The extraction of failure sequences from the polyacrylamide gel takes two days. Again, we did not optimize this process and extraction for a shorter time might work equally well. Cleavage of the oligo from the polyacrylamide gel including concentration of the extracts and oligo precipitation by *n*BuOH could take a day. PCR amplification of CBP-purified oligo including gel electrophoresis analysis of results can be accomplished in one day. All steps for cloning and sample preparation for Sanger sequencing can be finished in three days. Specifically, on day one, TOPO ligation can be accomplished first, and the recombinant DNA can be transformed into *E. coli* with overnight incubation. On day two, colony PCR can be performed, and positive colonies can be placed in liquid culture for overnight growth. On day three, plasmid DNA from overnight liquid culture can be isolated and prepared for Sanger sequencing. For the synthesis of PTP **2**, it will take one or two weeks to accomplish. The synthesis of phosphoramidite **3** takes one day.

CONFLICT OF INTEREST STATEMENT:

Michigan Technological University owns the IP (US7850949B2 and US9243023B2) of the catching-by-polymerization method.

DATA AVAILABILITY STATEMENT:

The Sanger sequencing data for oligo **1** are available at https://doi.org/10.26434/chemrxiv-2023-79qj3 or reference (Yin et al., 2023). See Supporting Information of the article.

ACKNOWLEDGMENTS:

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

LITERATURE CITED:

Azhayev, A., Gouzaev, A., Hovinen, J., Azhayeva, E., & Lonnberg, H. (1993). Analogs of oligonucleotides containing 3'-deoxy-beta-d-psicothymidine. *Tetrahedron Letters*, *34*, 6435-6438. doi:10.1016/0040-4039(93)85064-4.

Beaucage, S. L., & Caruthers, M. H. (2001). Synthetic strategies and parameters involved in the synthesis of oligodeoxyribonucleotides according to the phosphoramidite method. *Current Protocols in Nucleic Acid Chemistry*, doi:10.1002/0471142700.nc0471140303s0471142700.

Ceze, L., Nivala, J., & Strauss, K. (2019). Molecular digital data storage using DNA. *Nature Reviews Genetics*, *20*, 456-466. doi:10.1038/s41576-019-0125-3.

Chavda, V. P., Soni, S., Vora, L. K., Soni, S., Khadela, A., & Ajabiya, J. (2022). mRNA-based vaccines and therapeutics for COVID-19 and future pandemics. *Vaccines (Basel)*, *10*. doi:10.3390/vaccines10122150.

Eriyagama, D., Shahsavari, S., Halami, B., Lu, B. Y., Wei, F., & Fang, S. (2018). Parallel, large-scale, and long synthetic oligodeoxynucleotide purification using the catching full-length sequence by polymerization technique. *Organic Process Research & Development*, *22*, 1282-1288. doi:10.1021/acs.oprd.8b00209.

Eroshenko, N., Kosuri, S., Marblestone, A. H., Conway, N., & Church, G. M. (2012). Gene assembly from chip-synthesized oligonucleotides. *Current Protocols in Chemical Biology*, *2012*. doi:10.1002/9780470559277.ch110190.

Fang, S. Y., & Bergstrom, D. E. (2003). Fluoride-cleavable biotinylation phosphoramidite for 5'-end-labeling and affinity purification of synthetic oligonucleotides. *Nucleic Acids Research*, *31*, 708-715. doi:10.1093/nar/gkg130.

Fang, S. Y., & Bergstrom, D. E. (2004). Reversible 5'-end biotinylation and affinity purification of synthetic RNA. *Tetrahedron Letters*, *45*, 7987-7990. doi:10.1016/j.tetlet.2004.09.019.

Fang, S. Y., & Fueangfung, S. (2010). Scalable synthetic oligodeoxynucleotide purification with use of a catching by polymerization, washing, and releasing approach. *Organic Letters*, *12*, 3720-3723. doi:10.1021/ol101316g.

Garner, K. L. (2021). Principles of synthetic biology. *Essays in Biochemistry*, *65*, 791-811. doi:10.1042/EBC20200059.

Grajkowski, A., Cieslak, J., & Beaucage, S. L. (2016). Solid-phase purification of synthetic DNA sequences. *Journal of Organic Chemistry*, *81*, 6165-6175. doi:10.1021/acs.joc.6b01020.

He, M. H., Wu, X. S., Mao, S., Haruehanroengra, P., Khan, I., Sheng, J., & Royzen, M. (2021). Bio-orthogonal chemistry enables solid phase synthesis and HPLC and gel-free purification of long RNA oligonucleotides†. *Chemical Communications*, *57*, 4263-4266. doi:10.1039/d1cc00096a.

Hoose, A., Vellacott, R., Storch, M., Freemont, P. S., & Ryadnov, M. G. (2023). DNA synthesis technologies to close the gene writing gap. *Nature Reviews Chemistry*, *7*, 144-161. doi:10.1038/s41570-022-00456-9.

Kizerwetter, M., Pietz, K., Tomasovic, L. M., & Spangler, J. B. (2022). Empowering gene delivery with protein engineering platforms. *Gene therapy*. doi:10.1038/s41434-022-00379-6.

Ma, S., Tang, N., & Tian, J. (2012). DNA synthesis, assembly and applications in synthetic biology. *Current Opinion in Chemical Biology*, *16*, 260-267. doi:10.1016/j.cbpa.2012.05.001.

Masaki, Y., Onishi, Y., & Seio, K. (2022). Quantification of synthetic errors during chemical synthesis of DNA and its suppression by non-canonical nucleosides. *Scientific Reports*, *12*, 12095. doi:10.1038/s41598-022-16222-2.

Merritt, K. K., Bradley, K. M., Hutter, D., Matsuura, M. F., Rowold, D. J., & Benner, S. A. (2014). Autonomous assembly of synthetic oligonucleotides built from an expanded DNA alphabet. Total synthesis of a gene encoding kanamycin resistance. *Beilstein Journal of Organic Chemistry*, *10*, 2348-2360. doi:10.3762/bjoc.10.245.

Miklos, A. E., Hughes, R. A., & Ellington, A. D. (2012). Design and assembly of large synthetic DNA constructs. *Current Protocols in Molecular Biology*. doi:10.1002/0471142727.mb0323s99.

Pappas, K., Zhang, X., Tang, W., & Fang, S. Y. (2009). Phenyl esters, preferred reagents for mono-acylation of polyamines in the presence of water. *Tetrahedron Letters*, *50*, 5741-5743. doi:10.1016/j.tetlet.2009.07.142.

Pokharel, D., & Fang, S. Y. (2016). Polymerizable phosphoramidites with an acid-cleavable linker for eco-friendly synthetic oligodeoxynucleotide purification. *Green Chemistry*, *18*, 1125-1136. doi:10.1039/c5gc01762a.

Pon, R. T. (2000). Solid-phase supports for oligonucleotide synthesis. *Current Protocols for Nucleic Acid Chemistry*, Unit 3.1.1-3.1.28. doi:10.1002/0471142700.nc0301s00.

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. (2019). Optimizing genome assembly from pcr-amplified metagenomes. *Peerj*, *7*. doi:10.7717/peerj.6902.

Tang, W., & Fang, S. Y. (2008). Mono-acylation of symmetric diamines in the presence of water. *Tetrahedron Letters*, *49*, 6003-6006. doi:10.1016/j.tetlet.2008.07.174.

Tyagi, S., Kumar, R., Das, A., Won, S. Y., & Shukla, P. (2020). Crispr-cas9 system: A genome-editing tool with endless possibilities. *Journal of Biotechnology*, *319*, 36-53. doi:10.1016/j.jbiotec.2020.05.008.

Wincott, F. E. (2001). Strategies for oligoribonucleotide synthesis according to the phosphoramidite method. *Current Protocols in Nucleic Acid Chemistry*, doi:10.1002/0471142700.nc0471140305s0471142700.

Wurtzel, E. T., Vickers, C. E., Hanson, A. D., Millar, A. H., Cooper, M., Voss-Fels, K. P., Nikel, P. I., & Erb, T. J. (2019). Revolutionizing agriculture with synthetic biology. *Nature Plants*, *5*, 1207-1210. doi:10.1038/s41477-019-0539-0.

Yin, Y., Arneson, R., Apostle, A., Eriyagama, A., Chillar, K., Burke, E., Jahfetson, M., Yuan, Y., & Fang, S. (2023). Long oligodeoxynucleotides: Chemical synthesis, isolation via catching-by-polymerization, verification via sequencing, and gene expression demonstration. *Beilstein Journal of Organic Chemistry*, *19*, 1957-1965. doi:10.3762/bjoc.19.146.

FIGURES:

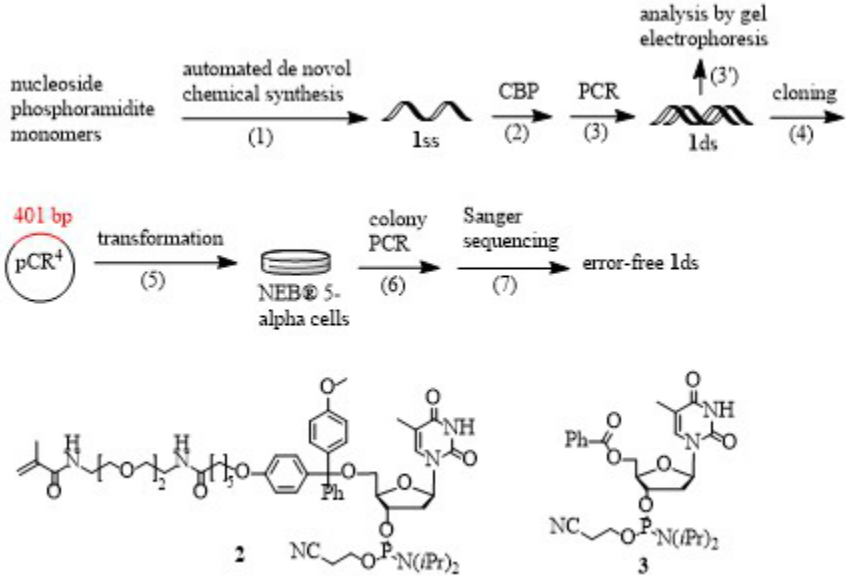

Figure 1. Workflow for producing error-free long oligos and the structure of phosphoramidites **2** and **3**. CBP, catching-by-polymerization. PCR, polymerase chain reaction. **1**ss, single-stranded oligo **1**. **1**ds, double-stranded oligo **1.** For nucleotide sequence of **1**ss, see Table 1.

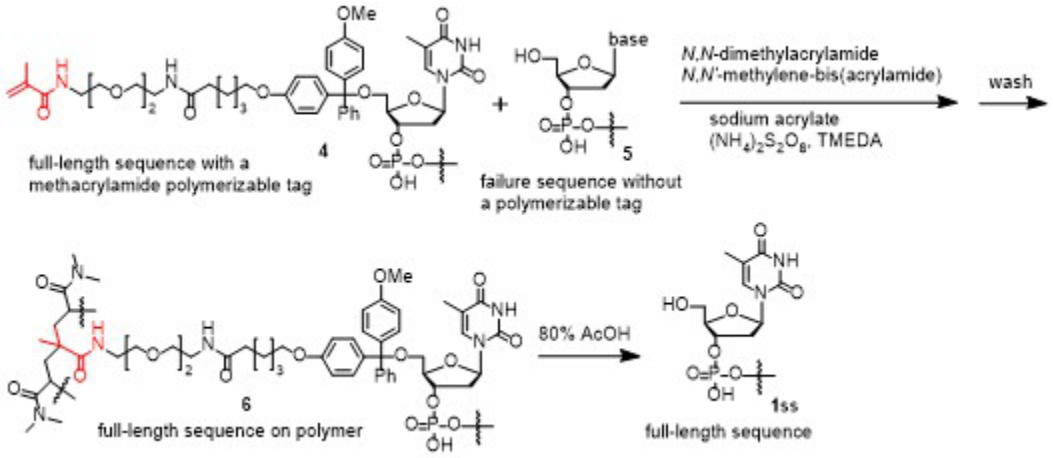

Figure 2. The catching-by-polymerization (CBP) process.

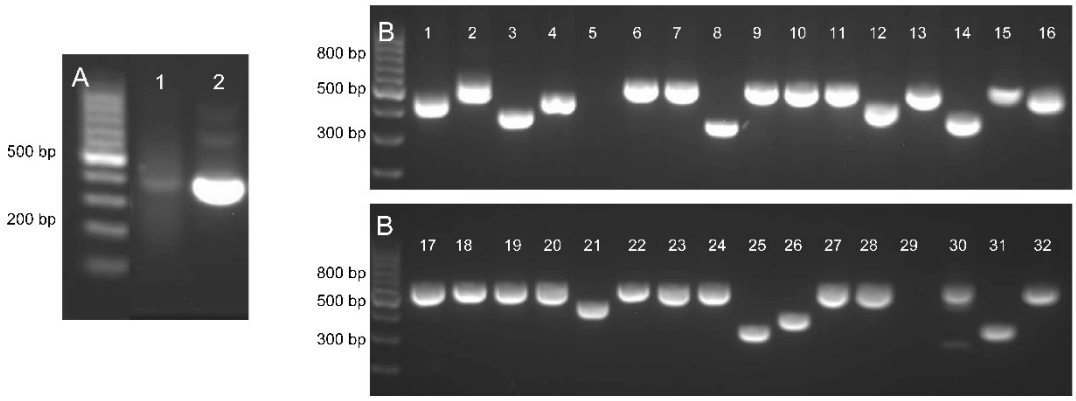

Figure 3. Gel electrophoresis images of **1**ds. (A) Lane 1: Image of **1**ds obtained by PCR with chemically synthesized and CBP purified **1**ss as template. Lane 2: Image of **1**ds obtained by PCR with commercial dsDNA as template. (B) Images of colony PCR products. Oligo **1**ds, used for lane 1 of (A), was cloned into a vector and transformed into *E. coli* cells. Thirty two colonies were selected for colony PCR using M13 primers. The products were analyzed with 2% agarose gel electrophoresis. Gel images were obtained by staining with GelGreen®. The M13 primers cover an additional 240 bp, and thus the PCR products containing **1**ds have about 641 bp, and those without **1**ds have 240 bp. Colonies corresponding to lanes 2, 6-7, 9-11, 13, 15, 17-20, 22-24, 27-28, 30 and 31 may have **1**ds. Colonies of lanes 6, 10 and 20 were submitted for Sanger sequencing; 6 and 20 had errors, 10 had the correct sequence. Total sequence alignment can be found online (Yin et al., 2023).

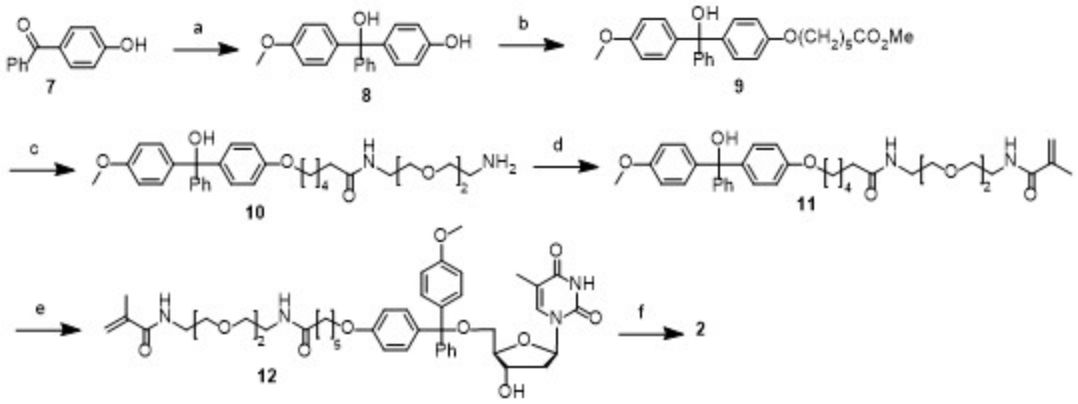

Figure 4. Synthesis of the polymerizable tagging phosphoramidites (PTPs) **2**. Conditions: (a) *p*-MeOPhMgBr (2.5 eq), THF, 0 °C to rt, 3 h, 84%; (b) $Br(CH_2)_5CO_2CH_3$ (1 eq), K_3PO_4 (2 eq), acetone, DMSO, reflux, 12 h, 77%; (c) $[CH_2O(CH_2)_2NH_2]_2$ (3 eq), H_2O, 90 °C, 12 h, 79%; (d) Methacrolyl chloride (1.1 eq), DIEA (5 eq), DCM, 0 °C to rt, 12 h, 90%; (e) TFAA (3 eq), DCM, 0 °C to rt, 2 h, remove volatiles, then thymidine (1.2 eq), DIEA (5 eq), DCM, pyridine, rt, 12 h, 77%; (f) $(i\text{Pr}_2\text{N})_2PO(CH_2)_2CN$ (1.1 eq), diisopropylammonium tetrazolide (1 eq), rt, 5 h, 82%.

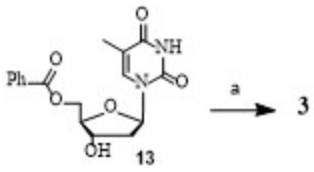

Figure 5. Synthesis of phosphoramidite **3**. (a) (iPr$_2$N)$_2$PO(CH$_2$)$_2$CN (1.5 eq), diisopropylammonium tetrazolide (1.5 eq), rt, 12 h, 89%.

TABLES:

Table 1. Sequences of oligo **1**ss and primers used in the protocols.

Name	Sequence
401-mer oligo (**1**ss)	(5')TATCACCTTCAAACTTGACTTCAGCACGCGTCTTGTAGTTCCCGTCATCTTTGAAAGATATAGTGCGT TCCTGTACATAACCTTCGGGCATGGCACTCTTGAAAAAGTCATGCCGTTTCATATGATCCGGATAACGG GAAAAGCATTGAACACCATAAGAGAAAGTAGTGACAAGTGTTGGCCATGGAACAGGTAGTTTTCCAGT AGTGCAAATAAATTTAAGGGTAAGTTTTCCGTATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAAT TTGTGCCCATTAACATCACCATCTAATTCAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTT TACTCATATTTTTTCCTCCTTATACTTAAGCCCTATAGTGAGTCGTATTAAT***T***CGC(3')
PCR forward primer (p**1a**)	(5')TATCACCTTCAAACTTGACTT(3')
PCR reverse primer (p**1b**)	(5')GCGAATTAATACGACTCACT(3')
M13 forward primer (p**1c**)	(5')GTAAAACGACGGCCAG(3')
M13 reverse primer (p**1d**)	(5')CAGGAAACAGCTATGAC(3')