Rapid and dynamic nucleic acid hybridization enables enzymatic oligonucleotide synthesis by cyclic reversible termination

or

A novel mechanism for enzymatic DNA synthesis

K. Hoff[1]*, M. Halpain[1], G. Garbagnati[1], J. Edwards[1, 2, 3, 4, 5], W. Zhou[1]

Affiliations:

[1]Department of Advanced Research and Development, Centrillion Biosciences, 2500 Faber Place, Palo Alto, CA 94303, USA.

[2]Chemistry and Chemical Biology, University of New Mexico, Albuquerque, NM 87131, USA.

[3]Internal Medicine, Chemical and Biological Engineering, University of New Mexico, Albuquerque, NM 87131, USA.

[4]University of New Mexico Comprehensive Cancer Center, Albuquerque, NM 87131, USA.

[5]University of New Mexico Health Sciences Center, Albuquerque, NM 87131, USA.

*Correspondence to: KHoff@Centrilliontech.com

Abstract:

Enzymatic oligonucleotide synthesis (EOS) has been attempted in many iterations for more than forty years, but chemical synthesis remains the industry standard despite hazardous waste produced, time restrictions, and length limitations of approximately 200 bases. Herein, we demonstrate that single-stranded oligos on a solid surface can transiently hybridize to neighboring strands and these structures can be recognized and extended by DNA polymerases and reverse transcriptases through a mechanism we describe as "bend and extend." Additionally, we show that the sequence of the newly synthesized fragment can be controlled to create custom oligonucleotides. We used this enzymatic approach to synthesize 20 bases on a solid surface through a two-step cyclic reversible termination process with stepwise efficiency over 98%. In our approach, a nascent DNA strand that serves as both primer and template is extended through polymerase-controlled sequential addition of 3'-reversibly blocked nucleotides followed by subsequent cleavage of the 3'-capping group. This process enables oligonucleotide synthesis in an environment not permitted by traditional phosphoramidite methods, eliminates the need for hazardous chemicals, has the potential to provide faster and higher yield results, and synthesizes DNA on a solid support with a free 3' end.

Introduction:

The first chemically synthesized dinucleotide, dTdT, was reported in 1955(1), and the first gene was synthesized in 1970(2). This gene, encoding a transfer RNA, was just 77 bases long and synthesized by enzymatically joining 17 chemically synthesized short oligonucleotides(2). While today, anyone can purchase custom oligonucleotide sequences up to several thousand bases long, the demand for longer and more accurate synthetic DNA is increasing as DNA and RNA are used for new applications in therapeutics(3,4), high-throughput genotyping(5), gene and whole genome synthesis(6), and data storage(7). These new applications place high requirements on stepwise yields, but chemical synthesis methods can only achieve 98.5-99.5% stepwise efficiency. The final yield for a 1kb fragment synthesized with 99.5% efficiency per step is less than 1%. Therefore, long sequences are enzymatically assembled from shorter, chemically synthesized fragments. They are then enzymatically processed and screened to remove inaccurate sequences or correct for errors. This process can be arduous and the ability to synthesize longer, more accurate sequences would enable emerging advanced technologies limited by this multi-step method(6).

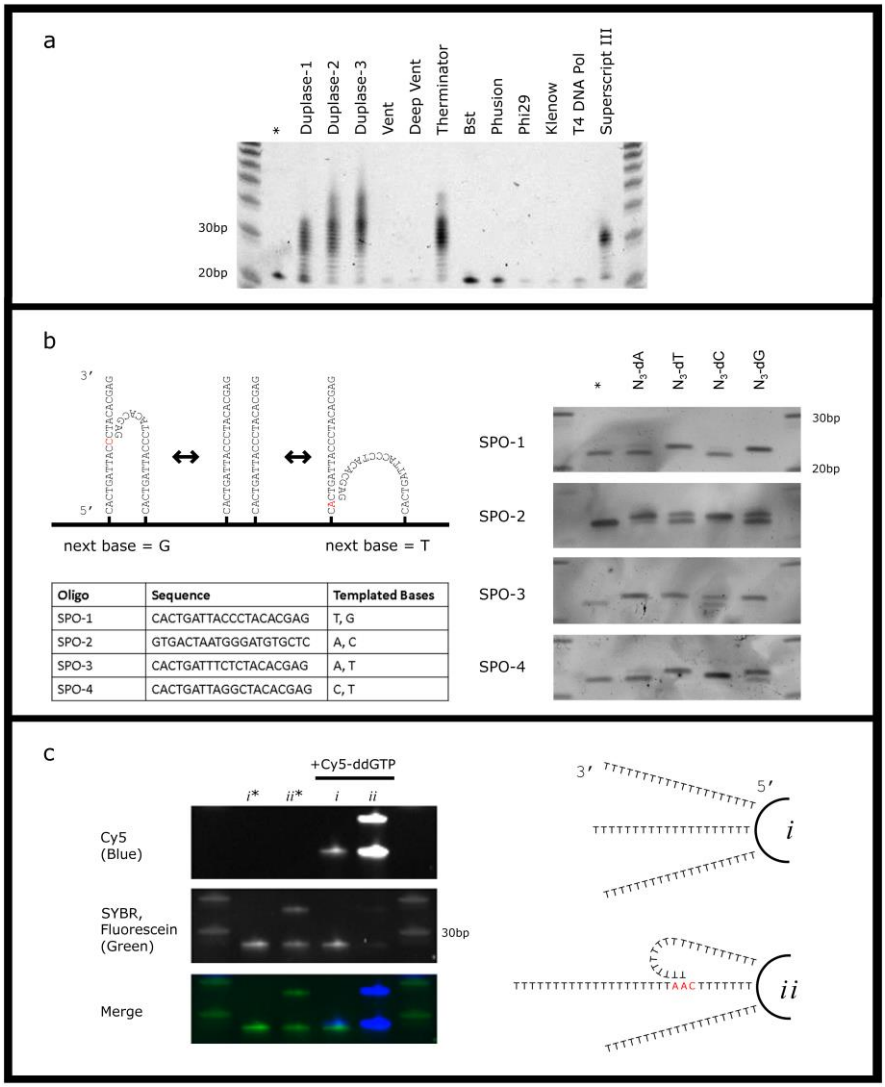
The first published Enzymatic Oligonucleotide Synthesis (EOS) methods used RNases(8) and polynucleotidyl phosphorylase(9). Oligos synthesized with these enzymes enabled research that was otherwise impossible(10), but the methods were highly inefficient compared with emerging phosphoramidite chemistry and subsequently abandoned. Attempts at EOS have also been made with ligases(11), but all recent published efforts have focused on Terminal deoxynucleotidyl Transferase (TdT). TdT is a unique DNA polymerase that can extend single-stranded DNA but not a DNA duplex due to a 16 amino acid lariat-like loop that occupies the binding site for a templating DNA strand(12). Several enzymatic oligonucleotide synthesis approaches using TdT have been attempted; cyclic reversible termination with 3'-blocked nucleotides(13–15), using a reversible covalent TdT-nucleotide complex(16), or simply using dNTPs to synthesize homopolymer tracts for data encryption(17,18). For various reasons, these methods have experienced limited success, and the longest product accurately synthesized using TdT is just 10 bases. However, as the demand for longer, more accurate, less expensive DNA increases, polymerase-mediated methods seem a natural direction to explore. In vivo, these enzymes are capable of high-fidelity synthesis of an entire genome(19). Therefore, we sought to develop a cyclic reversible termination method using a polymerase that efficiently incorporates 3'-blocked reversible terminators. Our method uses a polymerase to extend transiently hybridized structures in DNA, and herein we demonstrate controlled synthesis of single-stranded 20mer DNA oligonucleotide with stepwise efficiency over 98%. Ultimately, we envision a path forward using our methods to achieve efficiencies that exceed traditional phosphoramidite methods.

Results:

Transient Secondary Structures in DNA Enable Synthesis.

Some DNA polymerases and reverse transcriptases (RTs) are capable of extending solid-phase single-stranded DNA despite lacking nucleotidyltransferase activity (Fig. 1a). Therefore, we sought to harness this capability for EOS, and we generated a series of modified 9°N DNA polymerases (which is the basis for the Therminator polymerase [Fig. 1a]) capable of incorporating reversible terminators in these reactions, which we refer to as Duplases. The extension products from Duplases are templated single base extensions that are the result of transient hybridization, and Duplase extension proceeds in a sequence-specific manner with only two bases of transient hybridization are required for extension (Fig. 1b-c). Although extension with a nucleotide not templated by a neighboring sequence can be observed at longer timepoints, we hypothesize this is the result of misincorporation, and for shorter one-

minute reactions, complete incorporation with either of two templated bases is observed, and no incorporation is observed with the other bases (data not shown). We describe this mechanism as “bend and extend” synthesis.



Oligo	Sequence	Templated Bases
SPO-1	CACTGATTACCCTACACGAG	T, G
SPO-2	GTGACTAATGGGATGTGCTC	A, C
SPO-3	CACTGATTTCTCTACACGAG	A, T
SPO-4	CACTGATTAGGCTACACGAG	C, T

Fig. 1. Dynamic hybridization of DNA enables EOS. (a) Some polymerases and reverse transicrptases extend single-stranded DNA. Denaturing PAGE analysis of a 20 base single-stranded sequence, SPO-1, using different enzymes and dGTP. Despite a maximum of two bases of hybridization to a neighboring strand, some enzymes are able to extend this solid-phase oligonucleotide. None of these enzymes have previously been reported to have nucleotidyl transferase activity on single-stranded DNA. **(b) Sequence-specific extension of single-stranded DNA.** Denaturing PAGE analysis of four 20 base sequences in reactions with Duplase-3 and 3’-*O*-azidomethyl-dNTPs. Duplase extension of four different 20 base oligos proceeds in a sequence-specific manner, and only two bases of hybridization are required for extension. * indicates unextended control. Extension with non-templated bases can be explained by misincorporation that occurs with longer reaction times. **(c) Extension of solid-phase oligos through intermolecular reactions.** Magnetic beads were conjugated with either a 20 base poly-T oligo as illustrated in *i*, or with that oligo plus a 30 base poly-T oligo with an internal 5’-CAA-3’ sequence as illustrated in *ii*. Beads were extended using Duplase-3 and Cy5-ddGTP (* indicates unextended control during this step), and then all samples were labeled with TdT and fluorescein-12-ddUTP. Oligos appear blue if extended with Duplase-3 and green if extended wtih TdT.

The observed behavior of the Duplases supports a dynamic hybridization model as illustrated in Fig. 1b. Although it is widely understood that intramolecular reactions dominate in solution over intermolecular reactions(20), intermolecular reactions alone are sufficient to facilitate extension of surface-bound oligos (Fig. 1c). Poly-T sequences cannot efficiently hybridize with other poly-T sequences, so when the surface-bound oligo is a poly-T-mer, extension is extremely limited. However, robust extension is achieved when a second templating oligo is bound to the same surface and the poly-T-mer can hybridize to its neighbor even transiently. For complete extension of G on SPO-1, 100% of the oligos must at some point exist in the conformation where C is the templating base. For complete extension of T, 100% of the oligos must at some point exist in the conformation where A is the templating base. Because complete extension of each is observed, all oligos must pass through both conformations in one minute.

Polymerase-catalyzed cyclic reversible termination.

Some DNA polymerases are able to catalyze the incorporation of a reversible terminator on a solid phase oligo. The templating strand may be the nascent strand itself through hairpin formation and cis-extension, a neighboring strand, or another oligo provided in solution (Fig. 2). Oligos cannot be extended beyond a single base addition due to the 3'-OR blocking group on the newly added nucleotide. After the single incorporation event, the 3'-hydroxyl is restored through a cleavage event. This process may be repeated for the addition of a second, third, etc. base. Chemically-cleavable reversible terminators, 3'-*O*-azidomethyl-dNTPs, were used, and deprotection was achieved using Tris(2-carboxyethyl) phosphine hydrochloride (TCEP), as first described by Palla *et al.* (21). Other reversible terminators may be used, such as chemically cleavable 3'-aminoxy groups (22), enzymatically cleavable(23), or UV-cleavable(24,25) reversible nucleotides.

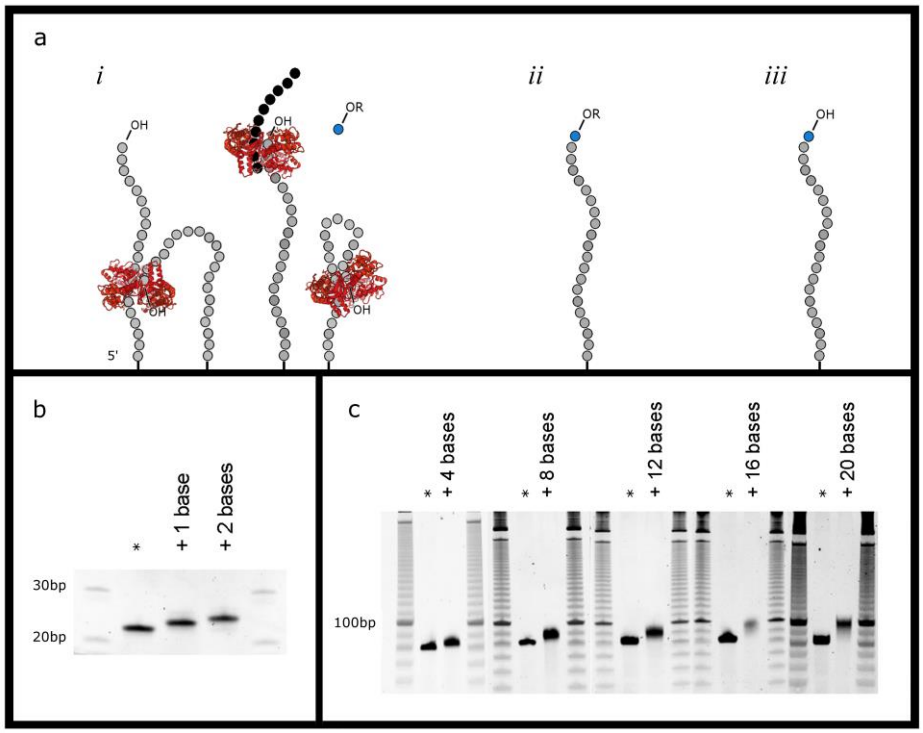


Fig. 2. (a) Polymerase-mediated solid-phase oligonucleotide synthesis using chemically blocked substrates. DNA/RNA polymerases and RTs require a double-stranded substrate to extend a nascent strand. (i) The templating strand may be a neighboring oligo, an oligo in solution, or the nascent strand itself in a self-priming hairpin structure. A 3'-reversibly blocked nucleotide is added by the enzyme in solution to produce the oligo depicted in (ii), one base longer, but not capable of being extended. The 3'-OR group is then converted into a 3'-OH group through a cleavage step producing the oligo in (iii), which can be extended again. **(b) Sequential incorporation of two bases using reversible terminators on a 20 base single-stranded oligo.** Denaturing PAGE analysis of a 20 base template (after incorporation of 3'-*O*-azidomethyl-dTTP (+1 base), followed by cleavage of the 3'-*O*-azidomethyl capping group and incorporation of 3'-*O*-azidomethyl-dCTP (+ 2 bases). * indicates unextended control. **(c) Synthesis of a 20 base single-stranded DNA fragment on a universal templating oligo.** Denaturing PAGE analysis after the addition of 4, 8, 12, 16, and 20 bases during the synthesis of the 20 base sequence, ESO-1, on a universal templating oligo using Duplase-3.

The enzymes.

Extension reactions using reversible terminators require enzymes whose catalytic domains are sufficiently large to accommodate the 3'-capping group. Many such enzymes have been reported, and numerous modified DNA polymerases have been developed for this purpose(26,27). Of the enzymes tested, the Moloney murine leukemia virus (M-MLV) RT, Superscript III, and the 9°N DNA polymerase, Therminator, are capable of extending solid-phase single stranded DNA with as few as two bases available for transient hybridization (Fig. 1a).

M-MLV RT and HIV-1 RT have been shown to possess a "clamping" ability requiring just two complementary bases, which enables SMART cDNA synthesis(28–30). SMART template switching requires non-templated addition

of nucleotides to the 3’ end of newly synthesized cDNA, which is maximally achieved by using manganese as a divalent ion cofactor(31). Though M-MLV RTs have previously been shown to incorporate some reversible terminators(32), the 9°N DNA polymerase has historically been the polymerase of choice in sequencing-by-synthesis experiments using the same 3’-O-azidomethyl reversible terminators used here(33,34). For this reason, we generated a series of modified 9°N DNA polymerases to develop optimized conditions for cyclic reversible termination. We refer to our enzyme as a Duplase (available from Centrillion Biosciences, please contact khoff@centrilliontech.com).

Comparative analysis of the fidelity of three Duplase enzymes is shown in Fig. 3. Traditional primer extension reactions show that Duplase-1 is higher fidelity than Duplase-2, which is in turn higher fidelity than Duplase-3. Decreasing enzyme fidelity increases the probability of misincorporation events, thus increasing the chance that a base may be added to a surface-bound oligo when the oligo may exist in a conformation that does not necessarily template the base in solution. The more promiscuous the enzyme, the higher the yield of the reaction product. Duplase-3 is used in this study for the controlled synthesis of the 20 base single-stranded sequence. Furthermore, Duplases are able to use ribonucleotides as a substrate (Supp. Fig. 1), and the 9°N polymerase has been used for XNA synthesis(35,36), increasing the options for substrates that can be synthesized using this technique.

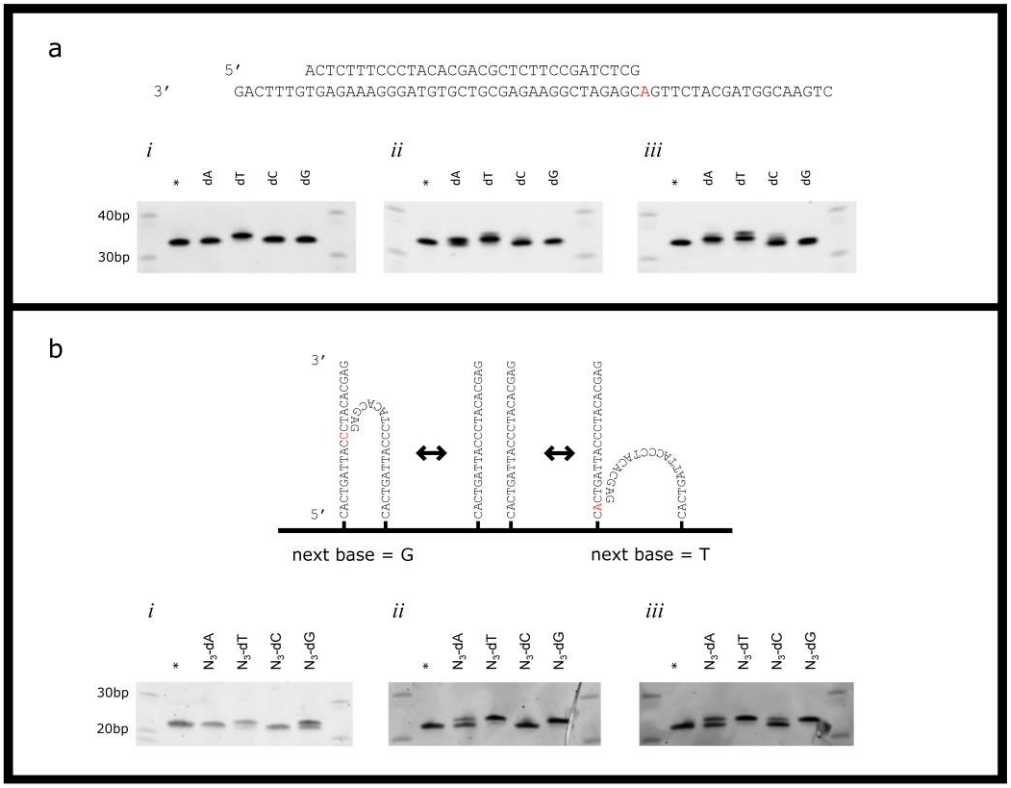


Fig.3. Comparison of the Duplase variants. (a) Fidelity on a double-stranded substrate. Primer extension reactions were conducted using the two oligos sequences shown with the longer templating strand bound to streptavidin-coated magnetic beads through a 5' biotin modification. The primer was stripped from the surface-bound template and analyzed by denaturing PAGE. Extension opposite adenine is shown with dATP, dTTP, dCTP, or dGTP used as the substrate in the reaction for (i) Duplase-1, (ii) Duplase-2, and (iii) Duplase-3. Decreasing fidelity can be seen. * indicates unextended primer control. **(b) Extension of a single-stranded substrate.** Reactions were conducted using a surface-bound single-stranded oligo as shown with one of four 3'-*O*-azidomethyl-dNTP in solution and (i) Duplase-1, (ii) Duplase-2, and (iii) Duplase-3. * indicates unextended primer control. As polymerase fidelity decreases, the capability to incorporate bases not templated by a neighboring strand or hairpin structure increases.

Optimization of reaction conditions.

We assayed variables to increase the chance of misincorporation and found that reaction temperature and the divalent metal ion used had the most significant impact. Manganese has been shown to alter the geometry of the polymerase's substrate binding pocket, opening it up to allow for polymerization with a non-templated base[3737], and polymerases are temperature sensitive with respect to fidelity(38). Optimal extension of single-stranded oligos was achieved using manganese (Supp. Fig. 2) at 60°C (Supp. Fig. 3). Because the lifetime of hybridization structures decreases with increasing temperatures, these data support a model of transient DNA hybridization where oligonucleotides on a solid surface are rapidly flipping between hybridization states that may be recognized and extended by a DNA polymerase or RT.

Extension opposite templates in solution.

Random hexamer priming is a well-established method for cDNA synthesis, rolling circle amplification, and multiple displacement amplification for whole genome amplification(39,40). Therefore, we sought to enhance the extension of a surface-bound oligo by the addition of in-solution 3'-blocked oligonucleotides. It was determined that randomers increase incorporation in reactions with Duplase-3 and a short solid-phase primer (Fig. 4). As with single-stranded oligos, reactions with random templates in solution are most efficient at higher temperatures, plateauing around 60°C (Supp. Fig. 3). Reaction yields are higher with shorter randomers in solution (Supp. Fig. 4).

When the templating strand is an oligo in solution as opposed to a neighboring strand, Duplases require three nucleotides of hybridization and at least one 3' base overhang for efficient solid-phase primer extension opposite an oligo in solution (Fig. 4). There are 1024 possible sequence combinations for pentamers, however, there are only four different possible sequence combinations exist for a semi-random sequence with a fixed 5' base, XNNNN where X templates the incoming base and N is a random base (A, T, C, or G). There are only 16 different possible sequence combinations for a semi-random sequence, XYNNN with two fixed 5' bases where X templates the incoming base and Y is complementary to the last base of the solid-phase oligo. For example, AGNNN should template the incorporation of thymine when the 3' base of the solid-phase primer is cytosine. Semi-random sequences increase the yield of extended primers (Fig. 4). However, there are obvious drawbacks to this method, for example when X=Y or for repetitive or homopolymeric stretches, and removal of templates from solution may be incomplete without additional purification steps and the use of templating oligos adds to the cost of the reactions; a universal templating oligo was instead used for the EOS of 20 bases as described below.

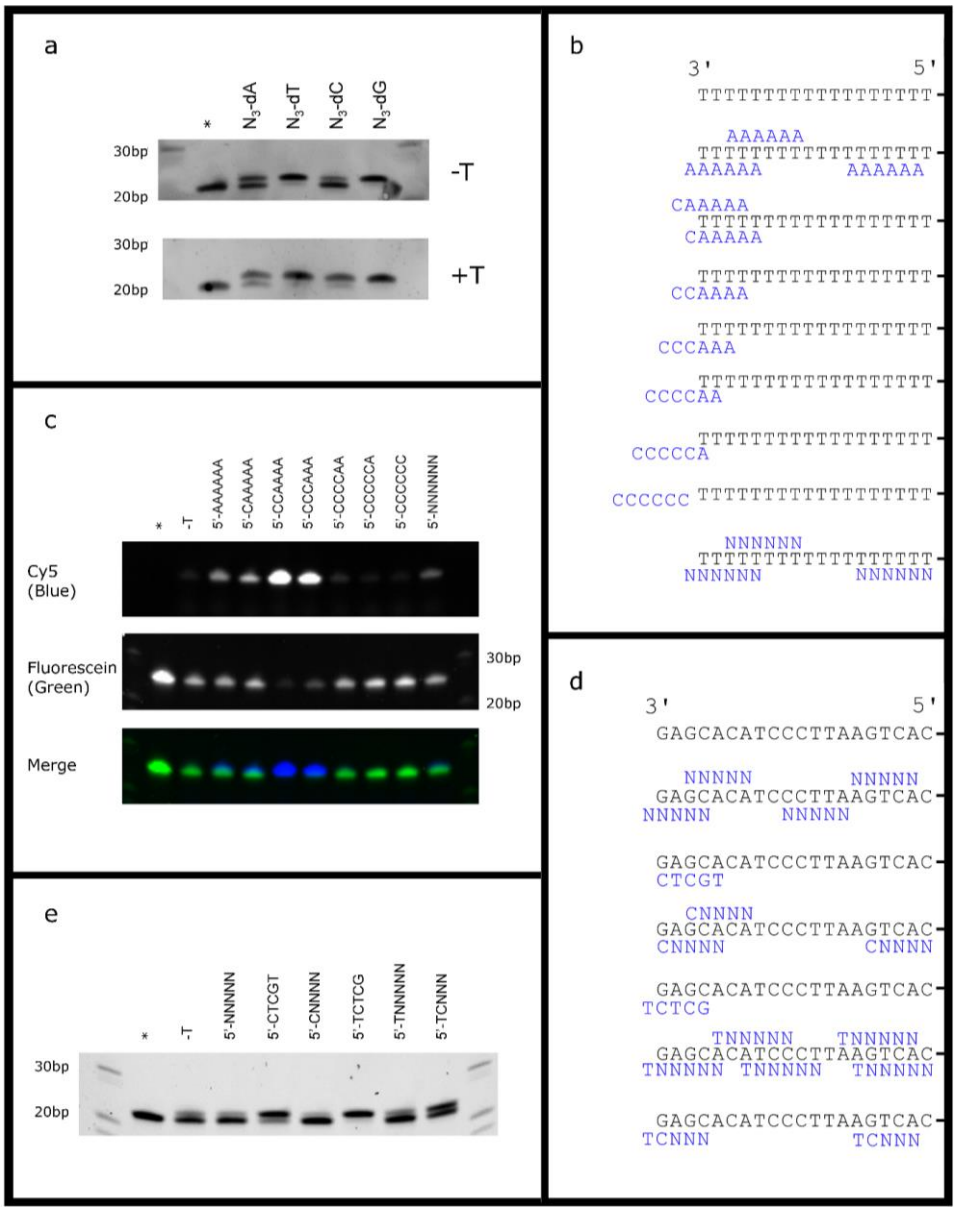


Fig. 4. (a) Extension of a short solid-phase oligo is enhanced by the addition of randomers to the reaction solution. Denaturing PAGE analysis of a 20 base oligo extended with Duplase-3 and 3'-*O*-azidomethyl-dNTPs without added template to solution (-T) or with the addition of 3'-phosphate blocked random hexamers to the reaction solution (+T). * indicates unreacted control. **(b) Possible hybridization of different templates to a solid-phase poly-T sequence.** Templates used in **(c)** are drawn in blue. **(c) Three bases of hybridization are required with a template in solution.** Denaturing PAGE analysis of a poly-T sequence extended with Duplase-3 and Cy5-ddGTP. 3'-phosphate blocked hexamers are used in reaction solutions with the sequences listed. * indicates unreacted control. -T indicates reactions without added template. After extension with Duplase-3, all samples were incubated with TdT and fluorescein-12-ddUTP to label any primer that was not extended with Duplase-3. **(d) Possible hybridization of different templates to a solid-phase oligo.** Templates used in **(e)** are drawn in blue. Some potential sequence alignments during transient hybridization reactions are illustrated. **(e) Extension of a solid-phase oligo with specific, random, and semi-random templates in solution**. Denaturing PAGE analysis of SPO-1 after extension with Duplase-3 and 3'-*O*-azidomethyl-dCTP. 3'-phosphate blocked pentamers are used in reaction solutions with the sequences listed. * indicates unreacted control. -T indicates reactions without added template.

Design of a universal template.

A universal templating oligonucleotide (UTO) templates any base regardless of the 3' sequence. UTOs were originally designed including the universal base, 5-nitro-1-indolyl-3'-deoxyribose (5-NI). The various properties of 5-NI have been extensively studied for over two decades since the original synthesis by Loakes and Brown(41–43). Unfortunately, there is no universal base that meets all desired requirements, ability to pair with all natural bases equally, prime DNA synthesis by DNA polymerases, and direct incorporation off each of the natural nucleotides(41,44). DNA polymerases do not efficiently incorporate nucleotides opposite or past 5-NI or any other universal base studied to date due to the lack of hydrogen bonding(45,46). Even when polymerases can overcome this barrier and incorporate opposite a universal base, they are not able to continue DNA synthesis past a position lacking hydrogen bonding, even under forcing conditions(41,47–49). We observed that Duplase extension of a poly-T sequence was not improved by consecutive internal 5-NI bases (data not shown).

UTO-1, a 78 base sequence (Supp. Table 1), contains all codons and includes poly-T sequences at the 3' and 5' ends that serve as spacers. The 3' deoxyuracil can be enzymatically cleaved using a combination of Uracil-DNA Glycosylase and apurinic/apyrimidic endonuclease to isolate any enzymatically synthesized fragment(50). All ANN, TNN, CNN, and GNN sequences are contained within UTO-1, meaning that any base can be templated during bend and extend EOS. We synthesized the 20 base sequence ESO-1 (enzymatically synthesized oligo-1) on UTO-1 using Duplase-3. The resulting oligo was poly-adenylated, and adapters were added by PCR (Supp. Fig. 5), and the library was sequenced on a MiSeq sequencer (Illumina, San Diego, CA).

Sequencing results for ESO-1.

83,631 sequences contained a partial UTO-1 sequence, UTO-1p (Supp. Table 1). These sequences were randomly divided into 10 groups for a statistical bootstrapped analysis of success rate, multiple base incorporation rate, and adjusted incorporation rate at each step of enzymatic synthesis (Fig. 5). The number of reads containing the full length ESO-1 sequence was compared to the number of reads one base shorter to determine the success rate for the perfect incorporation of the final base, and these calculations were repeated to determine synthesis success rate at each enzymatic synthesis step. Calculations excluded regions containing a homopolymer because when the n-1 base was the same as the n base or the n+1 base since failure to synthesize either base would result in the same partial sequence using these methods. A high percentage of sequences contained multiple bases added at a single step of

enzymatic synthesis, likely due to contamination of the nucleotides with 3'-hydroxyl groups (non-terminating bases) as seen by other researchers(17), which can be enzymatically corrected (17), or to TCEP carryover between steps as the rate of multiple incorporation for the same base varied at different synthesis steps. For example, some of the sequences containing NNNNNNNNNNNNNCAG read as NNNNNNNNNNNNNCAGG or NNNNNNNNNNNNNCAGGG. Incorporation of the final base, C, in NNNNNNNNNNNNNCAGC was thus calculated from the number of reads containing NNNNNNNNNNNNNCAG less the number of reads containing NNNNNNNNNNNNNCAGG.

Stepwise efficiency was evaluated this way for the last 12 bases of ESO-1 using the first 8 bases as a search parameter. This was repeated for a 14 base UTO-1 partial sequence, UTO-1p2 (Supp. Table 1), using the 19 preceding bases as a search parameter. This served as a control for comparison of sequencing results as it was replicated during PCR at the same time as ESO-1, but was originally synthesized by traditional phosphoramidite chemistry and subsequently HPLC purified. While the success rate mean for ESO-1 was only 87.9%, the incorporation rate mean adjusted for multiple base addition was 98.4%. For the control sequence, the success rate mean was 97.1%, and incorporation rate was 97.8% (Fig. 5). Errors arising during library preparation and sequencing account for the less than 100% rates we would expect for the control sequence.

a

$$^{5'}\text{NNNNNA}^{3'} \rightarrow {}^{5'}\text{NNNNNAC}^{3'}$$

$$\text{Success Rate} = \frac{\#\text{Reads(NNNNNAC)}}{\#\text{Reads(NNNNNA)}}$$

$$\text{Multi-base Rate} = \frac{\#\text{Reads(NNNNNACC)}}{\#\text{Reads(NNNNNAC)}}$$

$$\text{Incorporation Rate} = \frac{\#\text{Reads(NNNNNAC)}}{\#\text{Reads(NNNNNA)} - \#\text{Reads(NNNNNAA)}}$$

b

ESO-1

Sequence	Multi-base Sequence	Success Rate Mean	Success Rate SD	Multi-base Mean	Multi-base SD	Incorporation Rate Mean	Incorporation Rate SD
TACGATTCAGCCGATACAGC							
TACGATTCAGCCGATACAG	TACGATTCAGCCGATACAGG	0.97485	0.02592	0.01506	0.01957	0.98975	0.01681
TACGATTCAGCCGATACA	TACGATTCAGCCGATACAA	0.95160	0.04391	0.03812	0.04111	0.98933	0.01728
TACGATTCAGCCGATAC	TACGATTCAGCCGATACC	0.93391	0.03042	0.06001	0.03157	0.99365	0.01352
TACGATTCAGCCGATA*	TACGATTCAGCCGATAA	0.67789	0.04621	0.32211	0.04621	1.00000	0.00000
TACGATTCAGCCGAT	TACGATTCAGCCGATT	0.91956	0.03743	0.03741	0.01258	0.95530	0.03692
TACGATTCAGCCGA*	TACGATTCAGCCGAA	0.77241	0.05764	0.21970	0.05210	0.98945	0.01134
TACGATTCAGCCG	TACGATTCAGCCGG	0.67610	0.04572	0.31277	0.04710	0.98393	0.01541
TACGATTCAGCC+	TACGATTCAGCCC	0.75759	0.04390	0.22092	0.04504	0.97250	0.01221
TACGATTCAGC+	TACGATTCAGCC	0.94629	0.01263	0.94629	0.01263	18.51462	4.36685
TACGATTCAG	TACGATTCAGG	0.93428	0.01737	0.05126	0.01680	0.98478	0.00916
TACGATTCA	TACGATTCAA	0.86977	0.01804	0.11825	0.01368	0.98640	0.01234
Average		**0.87906**	**0.03456**	**0.10657**	**0.02931**	**0.98407**	**0.01660**

c

UTO-1 Control

Sequence	Multi-base Sequence	Success Rate Mean	Success Rate SD	Multi-base Mean	Multi-base SD	Incorporation Rate Mean	Incorporation Rate SD
CAAACTAATTTCAGAGTGACGTTACTTGGCGC							
CAAACTAATTTCAGAGTGACGTTACTTGGCG	CAAACTAATTTCAGAGTGACGTTACTTGGCGG	0.98933	0.00056	0.00273	0.00042	0.99204	0.00069
CAAACTAATTTCAGAGTGACGTTACTTGGC*	CAAACTAATTTCAGAGTGACGTTACTTGGCC	0.94259	0.00184	0.00768	0.00068	0.94988	0.00153
CAAACTAATTTCAGAGTGACGTTACTTGG*+	CAAACTAATTTCAGAGTGACGTTACTTGGG	0.99655	0.00082	0.00110	0.00040	0.99765	0.00061
CAAACTAATTTCAGAGTGACGTTACTTG+	CAAACTAATTTCAGAGTGACGTTACTTGG	0.99093	0.00090	0.99093	0.00090	110.24922	10.96611
CAAACTAATTTCAGAGTGACGTTACTT+	CAAACTAATTTCAGAGTGACGTTACTTT	0.98704	0.00071	0.00139	0.00031	0.98841	0.00077
CAAACTAATTTCAGAGTGACGTTACT+	CAAACTAATTTCAGAGTGACGTTACTT	0.98266	0.00095	0.98266	0.00095	56.83172	3.28747
CAAACTAATTTCAGAGTGACGTTAC	CAAACTAATTTCAGAGTGACGTTACC	0.94801	0.00191	0.00905	0.00073	0.95667	0.00168
CAAACTAATTTCAGAGTGACGTTA	CAAACTAATTTCAGAGTGACGTTAA	0.97182	0.00165	0.01888	0.00165	0.99052	0.00054
CAAACTAATTTCAGAGTGACGTT+	CAAACTAATTTCAGAGTGACGTTT	0.99128	0.00084	0.00181	0.00038	0.99308	0.00081
CAAACTAATTTCAGAGTGACGT+	CAAACTAATTTCAGAGTGACGTT	0.98281	0.00043	0.98281	0.00043	57.21700	1.41922
CAAACTAATTTCAGAGTGACG	CAAACTAATTTCAGAGTGACGG	0.98279	0.00155	0.00183	0.00035	0.98459	0.00151
CAAACTAATTTCAGAGTGAC	CAAACTAATTTCAGAGTGACC	0.92375	0.00286	0.01193	0.00056	0.93490	0.00263
CAAACTAATTTCAGAGTGA	CAAACTAATTTCAGAGTGAA	0.99063	0.00085	0.00362	0.00065	0.99423	0.00058
CAAACTAATTTCAGAGTG	CAAACTAATTTCAGAGTGG	0.99213	0.00111	0.00242	0.00090	0.99454	0.00067
Average		**0.97121**	**0.00150**	**0.00721**	**0.00075**	**0.97821**	**0.00119**

Fig. 5. Oligo Synthesis Efficiency. (a) Variables for evaluating efficiency. The success rate, multiple base incorporation rate, and adjusted incorporation rate are calculated for the extension of the oligo NNNNNA with D. These numbers are calculated for the bootstrapped sequences for ESO-1 **(b)**, and for the partial UTO-1 sequence used as a control **(c)**. * indicates oligos for which the n+2 base is the same so that the sequence NNNNNAA as in the example in part **(a)** may be due to multi-base extension or C not being incorporated. + indicates oligos for which the n+1 base is the same so that multi-base extension cannot be distinguished from accurate synthesis. These numbers were omitted from analysis.

As a secondary method of analyzing stepwise efficiency, alignment was evaluated for both sequences, as illustrated in Fig. 6, where each correctly matched base was scored with 1 and mismatches, insertions, and deletions were scored with 0. This allowed us to evaluate the length of our final fragment, accounting for multiple addition events. 54.6% of ESO-1 reads achieved the maximum score of 19 (the first base of the 20 base sequence was omitted because it is the same as the last base on UTO-1). 95.2% of the control sequence, UTO-1p2, reads achieved the

maximum score of 14. This results in an incorporation rate of 98.6% for ESO-1 (e^(ln(0.546)/19)) and 99.6% for UTO-1p2 (e^(ln(0.952)/14)).

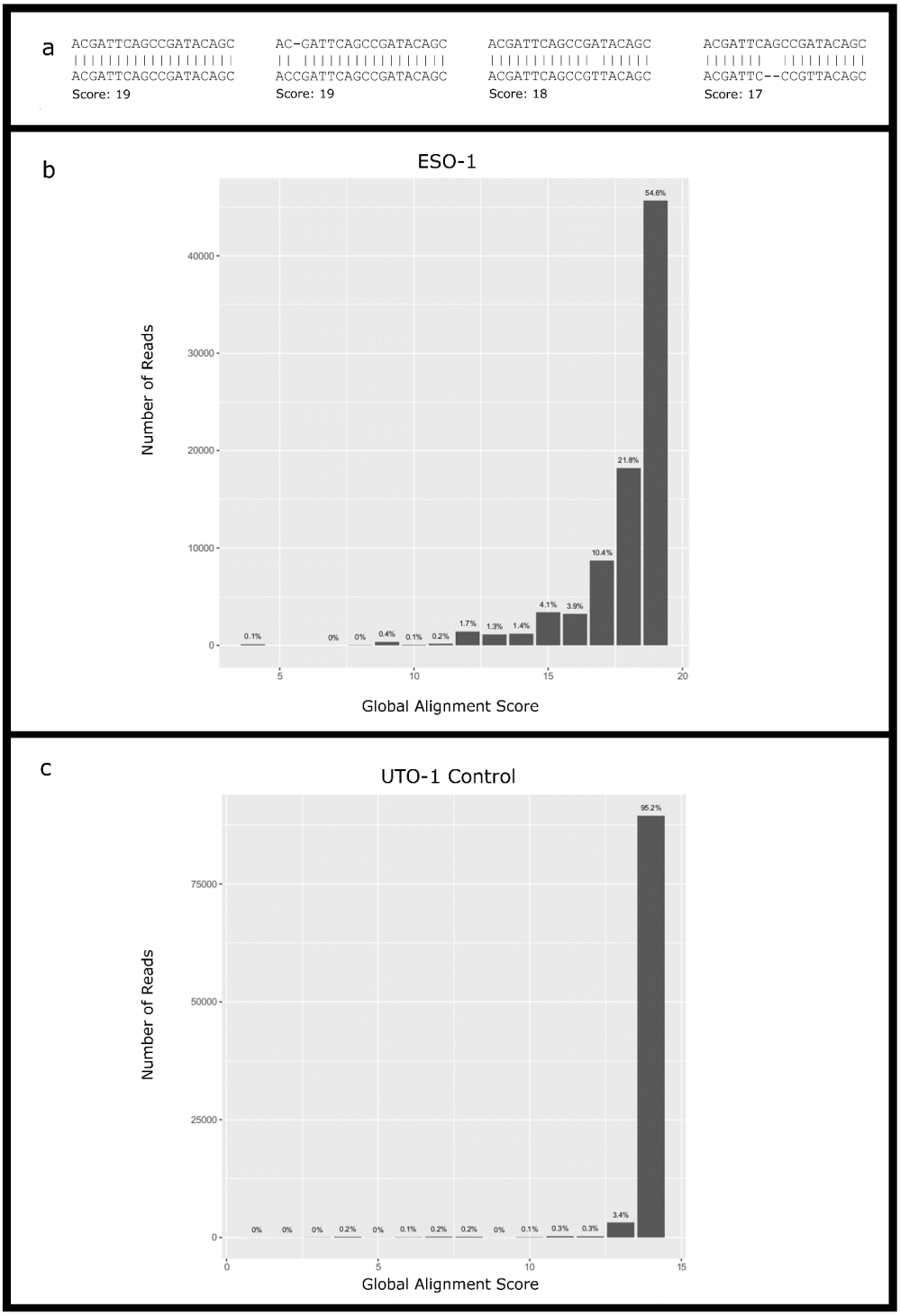


Fig. 6. Read 1 Alignment Scores. (a) Alignment Score Examples. A value of 1 is assigned to every perfectly matched base. A maximum value of 19 was possible for ESO-1 (one base was omitted from analysis because the

preceding sequence ended in the same base). A maximum value of 14 was possible for the PCR-prepared UTO-1p2 control sequence. Distribution of scores reflects final length of the correctly synthesized oligos accounting for multiple base incorporation, as displayed in **(b)** for ESO-1 and **(c)** for the UTO-1 control. 55.4% of sequences had a maximum value of 19 for ESO-1, and 95.2% of sequences had the maximum value of 14 for the UTO-1 control.

Discussion:

Since the isolation of the first DNA polymerase, *E. coli* pol I, by Kornburg in 1956, hundreds of other polymerases have been discovered, characterized, and engineered, and these polymerases have revolutionized the biomedical sciences with applications such as PCR and DNA sequencing. Herein, we have extended the capabilities of these enzymes and harnessed the same molecular machinery that made sequencing-by-synthesis possible to synthesize novel single-stranded DNA sequences. Enzymatic methods of DNA sequencing quickly overtook the Maxam-Gilbert chemical method more than 40 years ago; they were faster, less expensive, and able to produce longer reads. We hypothesize that DNA synthesis methods will follow a similar path with enzymatic methods opening the door for the production of longer, more accurate sequences at lower cost and without the production of hazardous waste.

We synthesized a 20 base single-stranded DNA oligo with stepwise yields between 98-99% using a DNA polymerase, reversible terminators, and a UTO. Our methods are already competitive with chemical methods, though further optimization will be required for the efficient synthesis of long oligos. Decreased enzymatic fidelity and mismatch discrimination would decrease reaction times and the length of the UTO required. Highly pure reversible terminators and a neutralization step following cleavage would eliminate multiple incorporation events. The yield of full-length products could thus be further increased by adding a capping step to terminate synthesis of unextended oligos after every step, and full-length products can be selected for by adding a final base either modified with a group that can be easily pulled-down such as biotin, or with one that may protect against exonucleolytic degradation in a cleanup step. It is reasonable to suspect that EOS can be improved to produce results greater than what is currently seen with chemical methods. High efficiency and low cost should encourage the use of this technology, leading to further advances.

Methods:

Primer extension assays on double-stranded DNA. Primers were hybridized to a biotinylated target sequence immobilized on streptavidin-coated magnetic beads (Dynabeads MyOne Streptavidin T1, Thermo Fisher, Waltham,

MA) by heating to 70°C for five minutes then 55°C for 15 minutes then 25°C for five minutes in RB (1M sodium chloride, 25mM Tris-HCl pH 7.5, 0.01% TWEEN20). 0.1mg of beads were used per 25µL reaction. Reactions were started by the addition of nucleotide at temperature and stopped by the addition of a 1.6x volume of RB. The hybridized primer was stripped from the bead-bound template using 0.1N sodium hydroxide for ten minutes at room temperature and analyzed by denaturing PAGE. All oligos were purchased from Integrated DNA Technologies (IDT, Redwood City, CA). Sequences are listed in Supp. Table 1.

Preparation of single-stranded DNA on a solid surface for extension.

Single-stranded oligos were immobilized to streptavidin coated beads. 0.1mg of beads were used per 25µL reaction volume. Reactions were started by the addition of nucleotide at temperature and stopped by the addition of a 1.6x volume of RB. The immobilized primer was stripped from the bead-bound template in 0.1N sodium hydroxide for five minutes at 65°C and analyzed by denaturing PAGE. Oligos were purchased from IDT and are listed in Supp. Table 1.

Primer extension reactions with Duplase enzymes.

For reactions on double-stranded DNA, magnetic beads were prepared as previously described with immobilized template and hybridized primer. Beads were washed in reaction buffer and then resuspended in the reaction mix containing a final concentration of 20mM Tris-HCl pH8.8, 10mM ammonium sulfate, 10mM KCl, 0.1% Triton X-100, 2mM $MgSO_4$, 1mg/mL bovine serum albumin (Sigma-Aldrich, St. Louis, MO), 4µg/mL polyvinylpyrrolidone 10, and 1ug of enzyme. Reaction mixtures were pre-warmed to 45°C, and nucleotide was added to a final concentration of 2µM. Reactions were allowed to proceed for one minute before stopping for analysis by denaturing PAGE. Extension reactions on single-stranded DNA were performed as previously described though buffer components, reaction times and temperatures, and nucleotide concentrations varied across experiments conducted for research purposes. Optimized conditions for the synthesis of the 20 base ESO-1 sequence are described below.

Extension reactions with commercially available enzymes.

Extension reactions on single-stranded DNA were performed similarly to the reactions with Duplase enzymes described above for 60 minutes at 60°C using 2U Vent (NEB, Ipswich, MA), 2U Deep Vent (NEB), or 2U

Therminator (NEB) in 20mM Tris-HCl pH8.8, 10mM ammonium sulfate, 10mM KCl, 0.1% Triton X-100, 8mM manganese chloride, 10uM dGTP. Extension reactions with 2U Phusion (NEB) were performed for 60 minutes at 60°C using Phusion high fidelity buffer, 8mM manganese chloride, 10uM dGTP. Extension reactions with 10U Phi29 (NEB) were performed for 60 minutes at 37°C using Phi29 buffer, 8mM manganese chloride, 10uM dGTP. Extension reactions with 5U Klenow (NEB) or 3U T4 DNA polymerase (NEB) were performed for 60 minutes at 37°C using NEBuffer 2, 8mM manganese chloride, 10uM dGTP. Extension reactions with 250U Superscript III (Thermo Fisher) were performed for 5 minutes at 25°C, 5 minutes at 50°C, then 30 minutes at 55°C using first strand buffer, 8mM manganese chloride, 10uM dGTP. Reactions were analyzed by denaturing PAGE.

Cleavage of Reversible Terminators: After incorporation of reversible terminators, beads were resuspended in 50mM TCEP pH9.0 (Gold Bio, Olivette, MO) and incubated at 60°C for 10 minutes. The TCEP solution was removed, and beads were resuspended in RB and transferred to a new reaction vessel.

Synthesis of the 20 base ESO-1 sequence with SPO-13.

Beads were prepared as previously described with immobilized SPO-13. 0.2mg of beads were used in 50µL reactions containing 200mM NaCl, 20mM Tris pH 8.0, 8mM $MnCl_2$, 1µg Duplase-3, and 100µM of a 3'-*O*-azidomethyl-dNTP. Reactions were allowed to proceed for one hour before stopping. The sequence ESO-1 was synthesized by the incorporation of 3'-*O*-azidomethyl-dCTP followed by cleavage with TCEP and the incorporation of 3'-*O*-azidomethyl-dGTP, followed by cleavage with TCEP, etc. For the first synthesis, after the addition of 4, 8, 12, 16, and 20 bases, 0.025mg of beads was removed from solution, and the oligo was stripped from beads and analyzed by denaturing PAGE. Volumes for next steps were adjusted accordingly. After the final incorporation/cleavage step, the oligo with ESO-1sequence was used for sequencing.

Fluorescent labeling of oligos with terminal deoxynucleotidyl transferase (TdT).

Labeling of poly-T sequences, which do not stain well with SYBR Gold, was accomplished by end-labeling oligos with TdT (New England Biolabs, Ipswich, MA) and fluorescein-12-ddUTP (Perkin Elmer, San Jose, CA) by incubating 0.1mg of beads with 5µM nucleotide and 20U of enzyme in 20mM Tris-HCl pH7.5, 10mM ammonium

sulfate, 10mM KCl, 0.1% Triton X-100, and 0.5mM $MnCl_2$. Reactions were allowed to proceed for 3 hours at 37°C before heat-inactivation of the enzyme for 20 minutes at 75°C.

Sequencing of SPO-13 with synthesized 20 base AM1 sequence.

The SPO-13-AM1 oligo still on beads was poly adenylated using TdT (Roche, Santa Clara, CA) and Duplase-1. Illumina sequencing adapters were added to these sequences using PCR, a poly(T)-tailed P7 adapter sequence, P7-Poly(T), and an oligo with the sequence PCR-ESO1-FPCR was repeated with Illumina P5 and P7 adapter oligos, and the library quality was analyzed by Bioanalyzer (Agilent, Sanat Clara) and qPCR. All oligos were purchased from IDT. A MiSeq nano flow cell (Illumina) was clustered with a 4pM library comprised of 90% PhiX and 10% PCR-prepared library. 150 step paired end cycling was run using a V2 kit (Illumina), and the AM1 sequence was identified by manually sorting first output reads.

Denaturing PAGE. Samples were resolved by electrophoresis in 6% or 15% polyacrylamide TBE-Urea gels (Thermo Fisher) in TBE. Gels were stained in SYBR Gold (Thermo Fisher) for five to ten minutes at room temperature and imaged on a ChemiDoc MP system (Bio-Rad, Hercules, CA). Fluorescent products were visualized before and after staining.

Acknowledgements:

We would like to thank Justin Costa and Janet Warrington for their help and expertise writing. We would also like to thank Foster Hoff for his help in designing UTO-1 and Jiayi Sun for her help in analyzing sequencing data.

Competing interest statement:

All authors are employees of and shareholders at Centrillion Biosciences. This research has been patented and may be commercialized in the future.

References:

1. Michelson AM, Todd AR. Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3′:5′-internucleotidic linkage. J Chem Soc Resumed. 1955 Jan 1;0(0):2632–8.

2. Agarwal KL, Büchi H, Caruthers MH, Gupta N, Khorana HG, Kleppe K, et al. Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. Nature. 1970 Jul;227(5253):27–34.

3. Goodchild J. Therapeutic oligonucleotides. In: Therapeutic Oligonucleotides. Springer; 2011. p. 1–15.

4. Dowdy SF. Overcoming cellular barriers for RNA therapeutics. Nat Biotechnol. 2017;35(3):222.

5. Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvänen A-C. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Res. 2000;10(7):1031–42.

6. Kosuri S, Church GM. Large-scale de novo DNA synthesis: technologies and applications. Nat Methods. 2014;11(5):499–507.

7. Organick L, Ang SD, Chen Y-J, Lopez R, Yekhanin S, Makarychev K, et al. Random access in large-scale DNA data storage. Nat Biotechnol. 2018;36(3):242.

8. Uchida T, Funayama-Machida C. Systematic Synthesis of Dinucleotides and Trinucleotides with RNases U2 N1 and a Non-Specific RNase from B. subtilis. J Biochem (Tokyo). 1977;81(5):1237–46.

9. GILLAM S, ROTTMAN F, JAHNKE P, SMITH M. Enzymatic synthesis of oligonucleotides of defined sequence: Synthesis of a segment of yeast iso-i-cytochrome c gene. 1977;5.

10. Smith M. Synthetic DNA and biology (Nobel lecture). Angew Chem Int Ed Engl. 1994;33(12):1214–21.

11. Schmitz C, Reetz MT. Solid-Phase Enzymatic Synthesis of Oligonucleotides. Org Lett. 1999 Dec 1;1(11):1729–31.

12. Motea EA, Berdis AJ. Terminal deoxynucleotidyl transferase: the story of a misguided DNA polymerase. Biochim Biophys Acta BBA-Proteins Proteomics. 2010;1804(5):1151–66.

13. Minhaz Ud-Dean SM. A theoretical model for template-free synthesis of long DNA sequence. Syst Synth Biol. 2008 Dec;2(3–4):67–73.

14. Mathews AS, Yang H, Montemagno C. 3′-O-Caged 2′-Deoxynucleoside Triphosphates for Light-Mediated, Enzyme-Catalyzed, Template-Independent DNA Synthesis. Curr Protoc Nucleic Acid Chem. 2017;71(1):13.17. 1-13.17. 38.

15. Mathews AS, Yang H, Montemagno C. Photo-cleavable nucleotides for primer free enzyme mediated DNA synthesis. Org Biomol Chem. 2016;14(35):8278–88.

16. Palluk S, Arlow DH, de Rond T, Barthel S, Kang JS, Bector R, et al. De novo DNA synthesis using polymerase-nucleotide conjugates. Nat Biotechnol. 2018;

17. Jensen MA, Davis RW. Template-Independent Enzymatic Oligonucleotide Synthesis (TiEOS): Its History, Prospects, and Challenges. Biochemistry. 2018;57(12):1821–32.

18. Lee HH, Kalhor R, Goela N, Bolot J, Church GM. Enzymatic DNA synthesis for digital information storage. bioRxiv. 2018;348987.

19. Blanco L, Bernad A, Lázaro JM, Martín G, Garmendia C, Salas M. Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. J Biol Chem. 1989;264(15):8935–40.

20. Jones WD, Feher FJ. Kinetics and thermodynamics of intra-and intermolecular carbon-hydrogen bond activation. J Am Chem Soc. 1985;107(3):620–31.

21. Palla M, Guo W, Shi S, Li Z, Wu J, Jockusch S, et al. DNA sequencing by synthesis using 3′-O-azidomethyl nucleotide reversible terminators and surface-enhanced Raman spectroscopic detection. RSC Adv. 2014;4(90):49342–6.

22. Hutter D, Kim M-J, Karalkar N, Leal NA, Chen F, Guggenheim E, et al. Labeled nucleoside triphosphates with reversibly terminating aminoalkoxyl groups. Nucleosides Nucleotides Nucleic Acids. 2010;29(11–12):879–95.

23. Zhou W, McGall G, Moti J, Mei R. Reversible terminator molecules and methods of their use. Google Patents; 2017.

24. Litosh VA, Wu W, Stupi BP, Wang J, Morris SE, Hersh MN, et al. Improved nucleotide selectivity and termination of 3′-OH unblocked reversible terminators by molecular tuning of 2-nitrobenzyl alkylated HOMedU triphosphates. Nucleic Acids Res. 2011;39(6):e39–e39.

25. Metzker ML, Raghavachari R, Richards S, Jacutin SE, Civitello A, Burgess K, et al. Termination of DNA synthesis by novel 3’-modifieddeoxyribonucleoside 5’-triphosphates. Nucleic Acids Res. 1994;22(20):4259–67.

26. Smith GP, Rigatti R, Ost TWB, Balasubramanian S, Sanches-Kuiper RM. Polymerases. Google Patents; 2015.

27. Smith GP, Bailey DMD, Sanches RM, Swerdlow H, Earnshaw DJ. Modified polymerases for improved incorporation of nucleotide analogues. Google Patents; 2013.

28. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. Reverse transcriptase template switching: A SMART™ approach for full-length cDNA library construction. Biotechniques. 2001;30(4):892–7.

29. Oz-Gleenberg I, Hizi A. Strand selections resulting from the combined template-independent DNA synthesis and clamp activities of HIV-1 reverse transcriptase. Biochem Biophys Res Commun. 2011;408(3):482–8.

30. Oz-Gleenberg I, Herschhorn A, Hizi A. Reverse transcriptases can clamp together nucleic acids strands with two complementary bases at their 3′-termini for initiating DNA synthesis. Nucleic Acids Res. 2010;39(3):1042–53.

31. Oz-Gleenberg I, Herzig E, Voronin N, Hizi A. Substrate variations that affect the nucleic acid clamp activity of reverse transcriptases. FEBS J. 2012;279(10):1894–903.

32. Knapp DC, Serva S, D’Onofrio J, Keller A, Lubys A, Kurg A, et al. Fluoride-Cleavable, Fluorescently Labelled Reversible Terminators: Synthesis and Use in Primer Extension. Chem- Eur J. 2011;17(10):2903–15.

33. Chen F, Dong M, Ge M, Zhu L, Ren L, Liu G, et al. The History and Advances of Reversible Terminators Used in New Generations of Sequencing Technology. Genomics Proteomics Bioinformatics. 2013 Feb;11(1):34–40.

34. Ju J, Kim DH, Bi L, Meng Q, Bai X, Li Z, et al. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. Proc Natl Acad Sci. 2006;103(52):19635–40.

35. Dunn MR, Larsen AC, Fahmi N, Zahurancik WJ, Meyers M, Suo Z, et al. Therminator-mediated synthesis of unbiased TNA polymers requires 7-deazaguanine to suppress GG mispairing during TNA transcription. J Am Chem Soc. 2015;137:4014–7.

36. Renders M, Emmerechts G, Rozenski J, Krecmerová M, Holý A, Herdewijn P. Enzymatic synthesis of phosphonomethyl oligonucleotides by therminator polymerase. Angew Chem. 2007;119(14):2553–6.

37. Zamft BM, Marblestone AH, Kording K, Schmidt D, Martin-Alarcon D, Tyo K, et al. Measuring cation dependent DNA polymerase fidelity landscapes by deep sequencing. PloS One. 2012;7(8):e43876.

38. Innis MA, Gelfand DH, Sninsky JJ, White TJ. PCR protocols: a guide to methods and applications. Academic press; 2012.

39. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, et al. Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci. 2002;99(8):5261–6.

40. Nelson JR, Cai YC, Giesler TL, Farchaus JW, Sundaram ST, Ortiz-Rivera M, et al. TempliPhi, phi29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. Biotechniques. 2002;32(6):S44.

41. Loakes D. Survey and summary: The applications of universal DNA base analogues. Nucleic Acids Res. 2001;29(12):2437–47.

42. Loakes D, Brown DM, Linde S, Hill F. 3-Nitropyrrole and 5-nitroindole as universal bases in primers for DNA sequencing and PCR. Nucleic Acids Res. 1995;23(13):2361–6.

43. Vallone PM, Benight AS. Melting studies of short DNA hairpins containing the universal base 5-nitroindole. Nucleic Acids Res. 1999;27(17):3589–96.

44. Loakes D, Hill F, Brown DM, Salisbury SA. Stability and structure of DNA oligonucleotides containing non-specific base analogues. J Mol Biol. 1997;270(3):426–35.

45. Berger M, Wu Y, Ogawa AK, McMinn DL, Schultz PG, Romesberg FE. Universal bases for hybridization, replication and chain termination. Nucleic Acids Res. 2000;28(15):2911–4.

46. Crey-Desbiolles C, Berthet N, Kotera M, Dumy P. Hybridization properties and enzymatic replication of oligonucleotides containing the photocleavable 7-nitroindole base analog. Nucleic Acids Res. 2005;33(5):1532–43.

47. Zhang X, Motea E, Lee I, Berdis AJ. Replication of a universal nucleobase provides unique insight into the role of entropy during DNA polymerization and pyrophosphorolysis. Biochemistry. 2010;49(14):3009–23.

48. Hoops GC, Zhang P, Johnson WT, Paul N, Bergstrom DE, Davisson VJ. Template directed incorporation of nucleotide mixtures using azole-nucleobase analogs. Nucleic Acids Res. 1997;25(24):4866–71.

49. Hill F, Loakes D, Brown DM. Polymerase recognition of synthetic oligodeoxyribonucleotides incorporating degenerate pyrimidine and purine bases. Proc Natl Acad Sci. 1998;95(8):4258–63.

50. Howell WM, Grundberg I, Faryna M, Landegren U, Nilsson M. Glycosylases and AP-cleaving enzymes as a general tool for probe-directed cleavage of ssDNA targets. Nucleic Acids Res. 2010;38(7):e99–e99.